DOCUMENT RESUME

ED 394 525 IR 017 828

AUTHOR	Miech, Edward J.; And Others
TITLE	On CALL: A Review of Computer-Assisted Language Learning in U.S. Colleges and Universities.
INSTITUTION	American Academy of Arts and Sciences, Boston, Mass.
PUB DATE	18 Apr 96
NOTE	115p.; From the Center for Evaluation of the Program Initiatives for Children.
PUB TYPE	Information Analyses (070) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Academic Achievement; Case Studies; Comprehension; *Computer Assisted Instruction; Educational Psychology; *Educational Research; *Educational Technology; Feedback; Higher Education; Linguistics; Literature Reviews; *Second Language Learning; Student Attitudes; Student Improvement
IDENTIFIERS	*Computer Assisted Language Learning; Empirical Research

ABSTRACT

This paper examines 22 empirical computer-assisted language learning (CALL) studies published between 1989 and 1994, and 13 reviews and syntheses published between 1987 and 1992, pertaining to CALL in higher education in the United States. A "three streams" framework helps to place CALL in a larger context and illustrate its several dimensions. Any specific CALL program involves decisions in relation to developments in at least three fields: educational psychology; linguistics; and computer technology. These three fields may be conceptualized as streams, where each stream flows more or less independently of the others, but where the practice of CALL at any given time requires making a passage across all three. An interpretive summary of five major findings from the review of the empirical CALL studies is offered: (1) captioning video segments can dramatically boost student comprehension; (2) CALL can connect students with other people inside and outside of the classroom, promoting natural and spontaneous communication in the target language; (3) the type of CALL feedback provided to students can play a central role in learning; (4) student attitudes toward CALL are not consistently linked to student achievement using CALL; and (5) CALL can substantially improve achievement as compared with traditional instruction. This paper also provides three general conclusions, each accompanied by recommendations for future CALL practice and research. Appendices include the material search procedure; captioning information; supplementary findings from the empirical studies; individual summaries of empirical studies; and individual summaries of CALL and Computer-Assisted Instruction (CAI) reviews. (Contains 43 references.) (Author/AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 394 525

On CALL:

A REVIEW OF COMPUTER-ASSISTED LANGUAGE LEARNING IN U.S. COLLEGES AND UNIVERSITIES

Edward J. Miech, Bill Nave, and Frederick Mosteller

April 18, 1996

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

From the Center for Evaluation of the Program on Initiatives for Children

of the

American Academy of Arts and Sciences, Cambridge, MA

Correspondence: Frederick Mosteller
Harvard University
Science Center, Room 603
1 Oxford Street
Cambridge, MA 02138

R017828

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Edward J. Miech
Bill Nave, Frederick Mosteller
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Table of Contents

Executive Summary 2

Acknowledgments 6

Definitions of Acronyms Used 7

Introduction: Computers in Education and Computer-Assisted Language Learning 8

Part One: Framework of Three Streams 11

Part Two: Major Findings from the Empirical Studies 14

Part Three: Review of CALL Reviews 36

Part Four: Recommendations for Future CALL Practice and Research 48

References 53

Appendix A: Search Procedure 61

Appendix B: Captioning Information 64

Appendix C: Supplementary Findings from the Empirical Studies 66

Appendix D: Individual Summaries of Empirical Studies 68

Appendix E: Individual Summaries of CALL and CAI Reviews 103

Executive Summary

Computers have now become part of the instructional landscape in formal educational settings from kindergarten through graduate school in the United States. As computer technology in education continues to proliferate, a three-decade-old question remains pertinent: how can computer technology be translated into improved teaching and learning? This question applies to the myriad subjects and contexts in which computer-assisted instruction has been implemented, including computer-assisted language learning (CALL) in U.S. colleges and universities, the focus of this report. In this paper, we consider 22 empirical CALL studies published between 1989 and 1994, and 13 reviews and syntheses published between 1987 and 1992, pertaining to computer-assisted language learning in higher education in the United States.

To avoid conveying an oversimplistic picture of CALL as merely a series of innovations in computer technology, a "three streams" framework helps to place CALL in a larger context and illustrate its several dimensions. Any specific CALL program involves decisions in relation to developments in at least three fields: (1) educational psychology, (2) linguistics, and (3) computer technology. These three fields may be conceptualized as three streams, where each stream flows more or less independently of the other two, but where the practice of CALL at any given time requires making a passage across all three.

This paper offers an interpretive summary of five major findings from our review of the empirical CALL studies:

1. Captioning video segments (providing on-screen subtitles in the target language) can dramatically boost student comprehension;

2. CALL can connect students with other people inside and outside of the classroom, promoting natural and spontaneous communication in the target language;

3. The type of CALL feedback provided to students can play a central role in student learning;

4. Student attitudes towards CALL are not consistently linked to student achievement using CALL; and

5. CALL can substantially improve student achievement as compared with traditional instruction.

This paper also provides three general conclusions, each accompanied by recommendations for future CALL practice and research:

1. Good CALL programs are hard to find because integrating the three streams of educational psychology, linguistics, and computer technology is difficult to do well.

Our recommendations:

- CALL developers could work in creative teams and combine different types of expertise to author and implement CALL programs.
- Educators in institutions of higher education could review CALL components of programs that have already been developed -- especially in federally-funded organizations -- and investigate the possibility of converting these programs for classroom use, as has been done with the Central Intelligence Agency's (CIA) Spanish-language *EXITO* program, which is now available to colleges and schools.

2. New multimedia computer technologies offer chances to develop more CALL programs that emphasize watching, listening, and speaking in addition to the traditional CALL activities of reading and typing, and new networking computer technologies provide opportunities to use CALL to promote person-to-person interaction in the target language that transcend obstacles of distance and time.

Our recommendations:

- Captioning video segments in the target language represents a way of leveraging multimedia computer technology into improved student foreign-language learning.
- The use of CALL as a vehicle for interpersonal communication with "distant others" in the target language over computer networks such as the Internet can promote content-rich conversations in which participants feel more invested in what they want to say than anxious over how correctly they say it.

3. The field of CALL needs more research, especially formative evaluation of CALL programs-in-development, conducted by a larger pool of researchers.

Our recommendations:

- More researchers in the field are needed because the responsibility for CALL research has fallen largely on the capable shoulders of a relatively small group of researchers plus a group of graduate students writing their theses on CALL.
- The CALL research literature might provide better guidance concerning how to use computers to improve teaching and learning in foreign languages in colleges and universities if CALL practitioners and researchers could agree upon a set of key questions for subsequent studies to address.

- Educators interested in implementing CALL programs on their own campuses could supplement the general insights offered in the CALL literature with formative research on programs-in-development to offer more context-specific findings.

Acknowledgments

Among those who have helped us in the course of this research project, George Conrad of Analysas advised us about the conversion of the *EXITO* program for classroom use. In addition, Kelly Nieves of George Mason University kindly made her doctoral dissertation available so that we could report on her preliminary research on her semester-long course based on *EXITO*. John Emerson of Middlebury College has advised us about materials and individuals who could respond to some of our concerns.

We would also like to thank Charles Abelmann, John Emerson, Nathan Keyfitz, Kelly Nieves, Lisa Soricone, Jane Tchaicha, and Cleo Youtz for their comments on an earlier version of this paper.

The Andrew W. Mellon Foundation has supported this work through a grant to the American Academy of Arts and Sciences.

DEFINITIONS OF ACRONYMS USED

CAI	Computer Assisted Instruction: Lessons mediated by a computer.
CALL	Computer Assisted Language Learning: Lessons in a second language mediated by a computer. The second language may be English for non-English speakers.
CD-ROM	Compact Disc - Read Only Memory: Data stored on a compact disc and readable by specialized computer hardware. Data may include text, sound, pictures, or video.
CIA	Central Intelligence Agency
ESL	English as a Second Language
L2	Second Language: The "foreign" language that a student is learning. The L2 for ESL students is English.
SLA	Second Language Acquisition: Related to Stephen Krashen's "Natural Approach," a theory that emphasizes the informal processes of acquiring a second language rather than the formal processes of language learning.

Introduction: Computers in Education and Computer-Assisted Language Learning

Computers have now become part of the instructional landscape in formal educational settings from kindergarten through graduate school in the United States. The commercial growth of microcomputers over the past decade, with computers becoming more powerful, more compact, and less expensive with each passing year, has been remarkable. The number of microcomputers in U.S. elementary and secondary schools jumped from 630,000 in 1985 to nearly 4 million in 1993, while the median number of K-12 students per computer plummeted from 42 to 14 during the same time period (*Statistical Abstract of the United States*, 1994, Tables 252 and 253). The number of undergraduates and graduate students who reported using computers at school rose from 30% in 1984 to 40% in 1989 according to Current Population Surveys conducted by the U.S. Bureau of the Census (*Digest of Educational Statistics*, 1993, Table 412), and one author writing on the subject has suggested that "it is not overly optimistic to estimate that virtually every institution of higher education in the United States has computers that are available to students, faculty, and administrators" (Ely, 1993, p. 53). Although nationwide statistics about computers in higher education are sketchier and less up-to-date than in elementary and secondary education, a 1994 survey of 435 U.S. colleges and universities indicated that 86% of these institutions had Internet network affiliations (up from 68% reported in the 1993 survey), 93% used CD-ROM technology, 19% had satellite uplink and downlink capabilities, and 10% had speech recognition technology; furthermore, about 30% reported that 100% of their students had access to electronic mail, and 37% provided student access to computer networks in dorm rooms (Munson, Richter, and Zastrocky, 1994, pp. 31, 44, 68-73, 121).

As computer technology in education continues to proliferate, a three-decade-old question remains pertinent: how to translate computer technology into improved teaching and learning. This question applies to the myriad subjects and contexts in which computer-assisted instruction (CAI) has been implemented, including foreign language teaching and learning in U.S. colleges and universities, the focus of this report. Advocates of computer-assisted language learning (CALL) in particular, and CAI in general, have long made many enthusiastic claims about the instructional power of computer technology in higher education. Empirical evidence, however, has rarely supported these assertions. Furthermore, the experiences of educators with previous technological innovations in education -- CALL, for example, inherited the mixed legacy of the audiocassette language lab, which largely did not meet the expectations of practitioners -- informs a healthy skepticism on the part of many towards computers. In light of the distinction between theory and practice concerning the use of computers to improve teaching and learning, our review of research on CALL provides useful information and insights that educators in colleges and universities may wish to consider when thinking about making a substantial investment of time, energy, and resources into computer-assisted language learning.

The CALL literature addresses a broad array of topics, including descriptions and reviews of particular software programs, accounts of innovative hardware configurations, theoretical considerations of the relationship between language acquisition processes and software design, discussions of teacher and student attitudes towards computers, reports on pilot CALL projects, and overviews of emergent computer technologies. The complete span of this literature is beyond the scope of this paper. Our review focuses on a subset: empirical studies and reviews that

evaluated various aspects of CALL through analyzing the differential learning outcomes of groups of students.

More specifically, we consider 22 empirical CALL studies published between 1989 and 1994, and 13 reviews and syntheses published between 1987 and 1992, pertaining to computer-assisted language learning in higher education in the United States. The cutoff date of 1989 for the 22 empirical studies was selected because of technological advances in CALL hardware and software in recent years, the relative paucity of empirical CALL studies in higher education in the United States published before 1989, and the discovery that several authors had already written a comprehensive summary of CALL research up to 1989 (e.g., Dunkel, 1991). The explicit set of inclusion criteria for the 22 empirical studies in this retrospective analysis, then, consisted of works: (1) on CALL, (2) in higher education, (3) in the United States, (4) published since 1989, (5) that considered the differential achievement of at least one group, (6) by analyzing at least one quantitative measure of student performance (see Appendix A for more information on the search procedure).

The 13 reviews in this retrospective analysis primarily focused on empirical studies that met the above criteria, with the exception that the studies considered in these reviews could have been published before 1989.

One general finding soon became apparent: CALL has no agreed-upon research agenda. Consequently, CALL researchers examine a wide variety of topics, only rarely giving in-depth consideration of any one particular subject. The diverse nature of this literature makes it hard to conduct comparisons across studies, difficult to support generalizations with empirical data, and impossible to carry out a meta-analysis. As a result, available evidence leads to few definitive statements about the efficacy of CALL in institutions of higher learning in the United States.

However, the 22 empirical studies and 13 reviews of CALL offer compelling insights into the conditions under which computer-assisted foreign language learning can work to improve student achievement. Our retrospective analysis provides an interpretive summary of these findings. Although these findings provide only partial answers to questions about the effective use of CALL in colleges and universities, this synthesis reaps some salient data-based conclusions from an extensive literature.

Part One: Framework of Three Streams

To avoid conveying an oversimplistic picture of CALL as merely a series of innovations in computer technology, CALL needs to be placed here into a larger context. A "three streams" framework helps to illustrate the different dimensions of CALL. Any specific CALL program involves decisions in relation to developments in at least three fields: (1) educational psychology, (2) linguistics, and (3) computer technology. These three fields may be conceptualized as three streams, where each stream flows independently of the other two, but where the practice of CALL at any given time requires making a passage across all three. A capsule description of each "stream" follows. (For a more exhaustive treatment of educational psychology, linguistics, or computer technology, see the comprehensive literature reviews in Avent, 1993; Fox, 1991; and Nieves, 1994).

The stream of educational psychology includes three major schools of thought: behaviorism, cognitive psychology, and humanistic psychology. Behaviorism, inspired by the work of B. F. Skinner in operant conditioning, emphasizes reinforcement of observable behavior through feedback and rewards (or punishment), and manifests itself in teaching and learning

through, among other methods, a stress on repeated drill and practice. Cognitive psychology, in direct contrast to behaviorism, concerns itself with the inner workings of the mind, and emphasizes the importance of meaning-making in learning. Humanistic psychology, associated with Abraham Maslow, emphasizes the subjective world of the individual, and emerges in education through areas such as concern for the attitudes, feelings, and learning styles of students. When different people design a CALL lesson on the same language concept using the same computer, they may come up with radically different CALL programs, depending on their preferred theories of learning and educational psychology. Computers themselves do not possess theories of learning; computer programmers and educators, consciously or unconsciously, bring those theories to the task.

The stream of linguistics includes structuralism, transformational grammar, and the Natural Approach. Structuralism focuses on the form and grammar of language, and appears in language learning through the direct translation method. Transformational grammar, which originated with Noam Chomsky, posits that humans have innate capacities for learning languages, and considers language learning to be a creative process assisted by intrinsic, universal discovery principles. The "Natural Approach," popularized by Stephen Krashen, emphasizes the informal acquisition of language and features key concepts such as the affective filter (the state of relative anxiety experienced by the language learner) and comprehensible input (the messages in the target language which are understandable to the language learner). As in the stream of educational psychology, computers do not subscribe to a theory of linguistics, and the learning experience associated with a particular CALL program relates to the linguistic hypotheses, as well as the preferred theories of learning, of the people who designed and implemented the program.

Finally, the stream of computer technology involves mainframe computers with "dumb" terminals, personal computers with autonomous capabilities, decentralized networks of personal computers and servers linked through cables and modems (e.g., local area networks, on-line services and databases, the Internet), and personal computers with enhanced capacities (e.g., increased random-access memory and hard-drive space, sound and video cards, CD-ROM drives). Throughout the 1960s and 1970s, the large, centralized mainframe computer was prominent, with primarily large organizations able to afford the high cost of developing software. Technology limitations translated into computer activities basically confined to reading and writing on a terminal. The explosive growth of personal computers during the 1980s contributed to the creation of computer labs in many schools and universities and permitted educators to design their own CALL programs with user-friendly authoring systems. The emergence of decentralized networks of personal computers and servers has allowed easy access to vast libraries of information distributed across large geographical areas as well as to authentic communication with other people not in the classroom or language lab; the rise of personal computers with enhanced capacities has facilitated high-quality audio and video interfaces that make it possible for language learners not only to read and write on the computer, but also to watch, listen, and speak in response to realistic situations. Again, as with the streams of educational psychology and linguistics, the place at which a CALL designer steps into the stream of computer technology has a strong bearing on the ultimate CALL teaching and learning experience.

Developments in the fields of educational psychology, linguistics, and computer technology proceed more or less independently of one another, but the three streams converge in one way or another in every CALL program. The large number of possible combinations from

these three streams, furthermore, harks back to our earlier observation about the diverse nature of CALL research. Although researchers in education can usually agree upon the definition of a variable like class size or college grade point average, and thus conduct logical cross-comparisons of studies on such subjects, various CALL programs may employ completely different uses of learning theory, linguistic approach, and computer technology, sharing little other than the general CALL designation.

Part Two: Major Findings from the 22 Empirical Studies

With this larger framework in mind, then, this retrospective analysis offers an interpretive summary of five major findings from a review of 22 empirical CALL studies published since 1989. (Other supplementary findings from these empirical studies can be found in Appendix C; more detailed information about each of the 22 studies can be found in Appendix D.)

1. Captioning video segments by including on-screen subtitles in the target language can dramatically boost student comprehension.

Captioning video segments used in foreign language instruction may be the most cost-effective measure a college or university can take to improve student learning. In captioning, lines of text appearing on the bottom of the screen provide a written account of the spoken dialogue in a video segment. One way to understand captioning is to imagine watching a foreign-language movie in French with subtitles, except that the subtitles are also in French and correspond exactly to what the characters say in the film.

The simultaneous presentation of language in spoken and written forms through captioning combines a branch of cognitive psychology called information-processing theory with

the Natural Approach: captioning provides more comprehensible input in the target language by engaging both a student's aural and visual sensory receptors. Pertinent computer technology ranges from simple video-cassette recorders (VCRs) linked to computers or television sets to sophisticated multimedia workstations.

Available software now makes the captioning process relatively affordable and straightforward to accomplish on a personal computer. For example, with two video-cassette recorders (VCRs), one personal computer, a video monitor, and a decoder, an individual can add captions in Spanish to a Spanish-language video segment by playing the original video on the first VCR, entering one block of text at a time (one to four lines of script prepared ahead of time using a standard word-processing program) by pressing "Enter" on the computer keyboard at the appropriate moment, and recording the captioned video on the second VCR.

In a 1991 article, Garza investigated the effect captions in the target language had on 70 students enrolled in intermediate/advanced English as a Second Language (ESL) and 40 students enrolled in an advanced Russian course. Students were randomly assigned into two groups--with captions and without captions--and students in both groups attended one-hour testing sessions where they viewed five "authentic" video segments in the target language. Students in the experimental group watched the video segments with captions, while students in the control group watched the same video segments without captions. For each segment, students were asked to answer ten multiple-choice questions written in the target language. Students were instructed to mark only answers for which they had a high degree of certainty, and to leave others blank. At the end of each testing session, five students were randomly selected to remain for a five-minute individual interview. In this interview, students were asked to retell one video

segment of their choosing, keeping as close as possible to original language of the segment. These interviews were tape-recorded, and their purpose was to determine if captions affect the way advanced students assimilate the inherent language of a video segment.

Students who watched the segments with captions had a mean gain of 75% in correct answers, a mean decrease of 61% in incorrect answers, and a mean decrease of 84% in unanswered questions over students who watched without captions. Average gains in correct responses were higher for Russian students (90%) than for English as a Second Language students (60%). Interviewed students who saw captioned segments consistently demonstrated greater ability to recall language of the video than students who did not see captions. Garza hypothesizes that "by adding the textual modeling of the captions, the essential language of the segment is made more accessible and, thus, (at least potentially) comprehensible to the learner" and concludes that "the most significant conclusion suggested by this study is that captioning may help teachers and students of a foreign language bridge the often sizable gap between the development of skills in reading comprehension and listening comprehension, the latter usually lagging significantly behind the former" (pp. 244, 246).

In her 1994 article, Borras studied the effects of captioning on the oral communicative performance of 44 students of intermediate French, where captioning was part of a multimedia CALL program. Students were randomly assigned to treatment groups with and without captions. As part of a multimedia program called *Practicing Spoken French,* students watched a video segment with or without captions, depending on their treatment group, and then answered comprehension questions about the video. Next, students wrote a draft about events they had

seen in the video, and then recorded in French an oral statement up to 3 minutes in length based on their draft.

Oral statements were scored using an assessment instrument developed by Borras that considers effectiveness, accuracy, organization, and fluency. Students in the groups with captions scored significantly higher than students in groups without captions on overall oral performance.

The Borras study suggests how multimedia computer technology has created a software bridge between captioning and CALL. Borras authored *Practicing Spoken French* using *Hypercard* and video editing software, which allowed her to integrate computers, video, and captions. As video segments become increasingly frequent components of multimedia CALL, captioning appears to be a worthwhile investment of resources. (See Appendix B for more information about computer software and hardware for adding captions to video).

Another inexpensive source of video with captions, particularly for educators involved with ESL, is closed-captioned television in the United States. Since July 1993, all new televisions sold in the U.S. come with built-in chips for decoding closed captions. With the press of a button, captions in English appear at the bottom of the screen for all closed-captioned programs at no cost to the viewer. Current closed-captioned television programming generally includes news programs, prime-time shows, major sporting events, children's shows, and Public Broadcasting System productions. Taping a television show with on-screen captions using a VCR captures the captions along with the image and sound. In addition, the same button on the TV set activates captions for over 10,000 captioned movies on videotape, which includes most new releases. Although this service was primarily designed for hearing-impaired individuals, ESL educators may find closed-captioned television to be a convenient source of captioned video for foreign language

instruction (although proprietary interests dictate that formal permission may be necessary to use this video in a CALL program). Furthermore, the United States and Canada share the same captioning format--called line-21 captioning--and French teachers with access to captioned television programming in French-speaking regions of Canada may want to consider this option. European countries use a captioning system called teletext that, due to its very fast transmission speed, cannot be recorded on a home VCR, and captioned programming is sparse in most other areas around the world (The Caption Center, 1995).

Whichever method educators use to to add captions to foreign language video -- and computer applications have simplified the process considerably -- evidence from these studies indicate that captioning can substantially improve student comprehension.

2. CALL can connect students with other people inside and outside of the classroom, promoting authentic communication in the target language.

Educators in colleges and universities now use CALL to engage students in conversations in the target language with other people both inside and outside of the classroom through local area networks and wider systems of networks such as the Internet. In a sense, this represents a logical progression in conceptualizing CALL. Early manifestations of CALL usually involved a closed relationship between the student and the terminal of a mainframe computer. This type of programmed, drill-and-practice instruction placed teachers in a largely peripheral role, as students interacted with the machine and could progress through the sequence of lessons alone. In the 1980s and 1990s, a new generation of CALL programs converted this "line" between the

endpoints of student and computer into a triangle, where the third point was a person -- a teacher, a tutor, or a fellow student -- actively involved in working with the student in the classroom or computer laboratory on the CALL lesson. In the 1990s, this triangle has been reconfigured into multidimensional networks where teachers use CALL to promote person-to-person interactions in the target language, often with "distant others" beyond the walls of the classroom, that transcend obstacles of distance and time.

This use of CALL as a vehicle for interpersonal communication in the target language relates most closely to the humanistic and cognitive currents in the stream of educational psychology and to the Natural Approach in the stream of linguistics. Interactions between people via computer tend to elicit individual, subjective perspectives on topics of mutual interest, and participants in these conversations usually focus on the content, or meaning, of language, rather than its form. The Natural Approach advocates this type of communication in the target language on the grounds that second language learning occurs most effectively when people feel more invested in what they want to say than anxious over how correctly they say it.

In a 1994 article, Dorothy Chun describes a two-semester study in which 15 students enrolled in her beginning German course engaged in up to 14 real-time class discussions on a local area network, with each discussion lasting about 20-25 minutes. Chun's entire section traveled to the computer laboratory to conduct these on-line discussions in German on topics Chun had announced earlier. During these discussions, participants typed comments and read what others wrote. Chun hypothesized that the different format for class discussions "would provide students with the opportunity to generate and initiate different kinds of discourse structures or speech acts" (p. 20).

Chun found that students averaged 8.4 entries per session, and that the ratio of simple sentences to complex sentences improved from 3 to 1 during the fall semester to 4 to 3 during the spring semester. Virtually every question posed by a student or by Chun during an on-line discussion received an answer, with the total number of replies (229) to Chun's questions numbering about twice as many as the total number of replies (126) to students' questions. The total number of student statements addressed to other students (198), added to the total number of questions asked by students (256), was greater than the total number of replies to questions (454), indicating to Chun that students interacted "directly with each other, as opposed to interacting mainly with the teacher" (p. 28). Chun concludes that the on-line class discussions helped the section move away from the traditional dynamic of teacher-centered interaction in the target language, as students were "definitely taking the initiative, constructing and expanding on topics, and taking a more active role in discourse management than is typically found in classroom discussion" (p. 28).

In their 1993 article, Terri Cononelos and Maurizio Olivia describe how students in an intermediate/advanced Italian class used Internet-based newsgroups and electronic mail (e-mail) to communicate with native speakers around the country and the world. Students selected a topic of personal interest pertaining to modern Italian culture, such as opera or women's rights, and investigated the subject through independent study. By the third week of the course, students had posted three messages each week on newsgroups located on the Internet, and had to respond to every reply they received at least once. The teacher checked students' contributions to the newsgroups for the quantity and quality of their writing. Students also responded to messages sent to them through e-mail, but the instructor did not monitor these responses. At the end of the

semester, students turned in a summary and analysis of their postings and the responses elicited by these postings. Cononelos and Olivia report that students received an average of three replies for each newsgroup posting they wrote, and that the participating students thought that both their confidence in using Italian and the quality of their writing in Italian improved as a result of the experience (pp. 530-531).

In a 1992 monograph, Francoise Hermann investigated a classroom where students had access to each other's written work through CALL. Hermann compared the performance of a section of students (n=11) enrolled in beginning French that used "agentive" CALL with another section of students (n=13) enrolled in the same course that used "instrumental" CALL. (In Hermann's study, "instrumental" refers to "using language for action" in socially meaningful tasks; "agentive" refers to "manipulating language," as in drill and practice.) Students were not randomly assigned to these two sections; instead, they enrolled in the different classes on a voluntary and informed basis. Hermann did not teach either section. Students in the "agentive" group used CALL to complete a series of nine fill-in-the-missing-word (cloze) sets of exercises in French based on the last eight chapters of the class workbook, whereas students in the "instrumental" group used CALL to create a classroom newspaper in French. The different versions of newspaper articles written by students in the "instrumental" group were stored on a shared computer directory that allowed students and the teacher to access all student work on the newspaper in various drafts, and students in this "instrumental" group also used electronic mail to send messages to each other, to their teacher, or to Hermann.

Because of the small sample size and non-equivalence between the "instrumental" and "agentive" groups in Hermann's study, significance tests comparing the performance of the

students in the two groups are inappropriate. Hermann's analysis of student mean scores on a battery of four pre- and post-tests during the first and last week of the two sections do suggest, however, that students in the "instrumental" CALL section did as well on these measures as students in the "agentive" CALL section. Hermann concludes, "The findings of this study indicate that an instrumental approach to the use of the computer in a first year, third quarter French as a foreign language class, and the changes it carries with it, is both an effective and workable alternative approach....Classes in foreign language education could consider using instrumental computer technology in contrast to the prevalent agentive modes of computer use" (p. 159).

The empirical studies of Chun, Cononelos and Olivia, and Hermann feature relatively small sample sizes, but their findings suggest the instructional merits of using computer networks imaginatively. Considered together, these studies indicate a promising direction for the future of CALL: educators can use computers as vehicles both to support new and different interaction among students and teachers in the target language and to create opportunities for students to converse with native speakers and others outside of the classroom and the university.

3. <u>The type of CALL feedback provided to students can play a central role in student learning</u>.

The feedback a CALL program gives in response to students' attempts to communicate in the target language, particularly when students make errors, can be of central importance. Types of CALL feedback range from the "wrong, try again" variety to detailed explanations of why the

answer was incorrect complete with examples of model sentences in which the language concepts in question appear in context.

The three streams (educational psychology, linguistics, and computer technology) encompass a wide variety of possible positions with respect to feedback and error correction. Within the stream of educational psychology, for example, behaviorist principles generally follow a "zero tolerance" approach, in which student errors are immediately corrected lest the student mistakenly internalize the wrong ideas and later have to "unlearn" these misconceptions. In contrast, the stream of linguistics includes the Natural Approach, which recommends a more lenient approach towards error correction, as this approach believes that too much emphasis on the formal rules of the target language can interrupt students' tentative attempts to communicate in a new language, raise students' anxiety about language learning (i.e., clog the "affective filter"), and impede the process of second language acquisition. The stream of computer technology can affect the selection of an error correction strategy in CALL insofar as more sophisticated feedback requires software and hardware configurations capable of supporting artificial intelligence.

In a study published in 1993, Noriko Nagata randomly assigned 34 college students enrolled in an intermediate Japanese course to two groups-- a group that used a CALL program that provided conventional feedback on a lesson involving the construction of passive sentences and another group that used an "intelligent" CALL program on the same subject that gave detailed error analysis -- and compared their performance. The CALL program offering "conventional feedback" gave information in English about *what* was wrong with a student's answer in Japanese after comparing the student response with the correct answer stored in the

computer, whereas the "intelligent" CALL program explained in English *why* a student response was incorrect through employing artificial intelligence. Nagata demonstrates the difference between the two feedback strategies with an example of the different messages the CALL programs would give to students making the same error in Japanese:

> For this response, T-CALI ["traditional" CALL] provided this feedback: "GA is not expected to be used here. NI is missing. MOMAREMASU is wrong." I-CALI ["intelligent" CALL], however, provided not only these messages but also more detailed grammatical explanations about the errors, e.g., "in your sentence, GAKUSEE is the 'subject' of the passive (the one that is affected by the action), but it should be the 'agent' of the passive (the one who performs the action and affects the subject). Use the particle NI to mark it. The predicate you typed is in the imperfective form. Change it to perfective. Since you are talking with your friend and your friend is using the direct-style (casual style), use the direct-style for your response." (Nagata, p. 335)

The students participating in the study spent about four hours studying their respective CALL lessons, and did not know that a comparison of the two different types of feedback was being conducted. Nagata found that the students in the "intelligent" CALL group significantly outscored the students in the "traditional" CALL group on both a 20-question achievement test on passive sentence construction administered shortly after the last CALL session and a series of four questions pertaining to passive sentences on the final exam administered three weeks later. Nagata concludes that "the study reveals that the students had difficulty learning Japanese particles, and that the intelligent CALI [CALL] feedback, which explained the functions and semantic relations of nominal phrases in the sentence, was especially helpful to them for understanding the concepts of the particles and passive structures" (p. 337).

In a 1992 article, Bernadin Bationo investigated differences among various types of traditional feedback. Bationo randomly assigned 56 students enrolled in beginning French into four CALL groups receiving either written feedback, spoken feedback, written and spoken feedback combined, or no feedback, where all feedback was given on a Macintosh computer in English. Students in the four groups used the same CALL tutorial to study four lessons on the future indicative mood of regular verbs, receiving the feedback specified for their group.

Bationo found that students receiving written and spoken feedback combined outscored the students in the other three groups on the immediate post-test, with significant differences between the group with written and spoken feedback combined and the groups with written feedback and no feedback (see Table 1). The differences among the four groups were not

Table 1. Comparison of mean scores of students (n=56) in four feedback groups on pre-test, immediate post-test, and delayed post-test.

	Written & Spoken	Spoken	Written	None	F	p
pre-test	0.5	1	0	0.5		
immediate post-test	18.1	14.7	11.5	11.6	4.03	.01
delayed post-test	15.3	12.1	10.6	14.9	2.00	.12

significant for the delayed post-test administered to the students two days later. Bationo notes that the mean score for the no feedback group was surprisingly high on the delayed post-test (14.9), and conjectures that students in the group might have been so frustrated about not receiving any information about their mistakes on the CALL tutorial that they studied the material on their own outside of class (p. 51). Bationo concludes by suggesting that the students in the

group with written and spoken feedback combined performed the best because the simultaneous delivery of visual and oral information was most suitable for students with various learning styles and abilities (pp. 47, 51).

In sum, the findings from these studies demonstrate the importance of paying attention to the type of feedback offered by CALL programs, as different feedback strategies can result in different learning outcomes for students.

4. No apparent relationship consistently links student attitudes towards CALL with student achievement using CALL.

The finding that student attitudes towards CALL do not relate consistently to student achievement using CALL surprised the authors of several studies who collected both student test scores and self-reported survey data. These authors posited a hypothesis that sounded reasonable at the outset of their studies: CALL would be more effective with students who reported positive attitudes towards CALL, and less effective with students who reported negative attitudes. This hypothesis, however, was not supported by the evidence.

In a 1989 article on a CALL program for students enrolled in beginning French, Robert Fischer performed correlational analysis of student attitudes towards various components of the CALL program, and post-test achievement scores on those same components: vocabulary, discrete-point grammar, integrated grammar, and irregular verb morphology (see Table 2). The only statistically significant correlation between student test scores and student ratings of the usefulness of particular CALL exercises was for vocabulary items ($r = .623$, $p<.001$), and Fischer

Table 2. Correlations between student ratings (n=34) and post-test scores on four components of a CALL program.

	vocabulary	discrete-point grammar	integrated grammar	irregular verbs
student ratings	.623	.284	.292	.241

hypothesizes that this was because this vocabulary was not taught during classroom instruction. Fischer concludes, "The lack of clear relationships between students' perceptions of these CALL lessons and their relevant posttest scores indicates that they did not generally perceive the instructional value of the lessons directly in terms of their end-of-semester achievement" (p. 88).

In a 1992 study about the use of CALL to improve the English pronunciation of a group of international teaching assistants, Stenson et al. reported that the 18 participants in the experimental section and their tutors expressed great enthusiasm for the CALL program, but that this enthusiasm did not translate into superior performance: "The fact that the quantitative results do not show more than very minor differences between the experimental and control groups, while the qualitative results suggest that instructors and ITAs [international teaching assistants] alike were enthusiastic about the use of SpeechViewer, is problematic" (p. 14).

In a 1993 article, Jing-Fong Hsu, Carol Chapelle, and Ann Thompson investigated how student exploration within a CALL program correlated with student attitudes for 34 students enrolled in intermediate and advanced English as a Second Language courses at Iowa State University who participated in the study. The authors reported that there were no significant correlations between exploration -- operationalized as the number of sentences constructed by students during their four hours using the CALL program -- and student attitudes towards

computers, learning English, and the specific CALL program used in the study, and that the correlation with CALL in general (.25) was significant ($p<.05$) but weak (see Table 3).

Table 3. Correlations between mean number of sentences constructed by students (n=34) and student attitudes towards computers, learning English, CALL in general, and the specific CALL program used in the study.

	ATTITUDES TOWARDS			
	computers	learning English	CALL	CALL program
mean number of sentences constructed by students when exploring CALL	.16	-.09	.25	.006

Hsu, Chapelle, and Thompson state that "it was anticipated that students' attitudes would be correlated with their amounts of exploration" (p. 13) in their conclusion, but these expected relationships did not surface in the overall correlations.

The overall finding from these studies that student attitudes towards CALL are not consistently linked to student achievement using CALL demonstrates the need for formal measures of learning when assessing the effectiveness of CALL, as students' favorable or unfavorable opinions of CALL do not appear to translate directly into how much they gain from computer-assisted language learning.

5. CALL can substantially improve student achievement, as compared with traditional instruction.

Although students using a given CALL program will not always outperform an equivalent group of students in a traditional college foreign language course, at least one well-designed study

has documented a situation in which students in experimental sections of CALL markedly outscored students in control groups on measures of foreign language achievement (Avent, 1993). The evidence from this study establishes that participation in a CALL program has the potential to improve student achievement in a foreign language.

Of the 22 empirical studies we reviewed, only 4 directly compared CALL instruction with non-CALL instruction; it was much more common for studies to compare one type of CALL program with another type of CALL program. In addition to the Avent study, 2 other studies (Nieves, 1994; Wright, 1992) found significant gains for students using CALL as compared with students in traditional classrooms, while a third study that focused on the use of CALL for pronunciation training (Stenson, 1992) did not.

In the study reported in his 1993 dissertation, Avent recruited a volunteer pool of 272 students enrolled in beginning German. Students were placed into one of three "achievement level" groups (low, middle, or high) based on their course grades in the previous German course, and then randomly assigned to the CALL or control group.

Instead of going to the language laboratory like the students in the control group, students in the experimental group went to a Macintosh computer lab and worked through the German courseware designed by Avent. Avent reports that the CALL lessons took him approximately 250 to 300 man-hours to develop and test. This CALL courseware covered four units, with each unit including one program focusing on vocabulary and a second program focusing on grammar. Students were required to answer at least 80% of items correctly in exercises and achievement checks before proceeding to the next part. Students in the experimental group spent an average of nearly six hours in the computer lab, while students in the control group spent an average of four

hours in the language lab. Other than these hours spent in the computer and language labs, students in both groups learned German through traditional classroom instruction during this one-quarter course.

The main evaluation instrument was the final exam, which consisted of a section on listening comprehension and a section on grammar that offered a direct comparison of achievement between students in the CALL and those in the control groups. An additional vocabulary test was administered at the end of the quarter which allowed Avent to look at just the experimental group and perform a within-group analysis to compare students' understanding of vocabulary words taught through CALL versus vocabulary words taught by traditional methods over the semester (e.g.,, oral and written review in the classroom).

The mean score on the final exam among students in the CALL group was higher for grammar (Table 4) and vocabulary (Table 5) test items than the mean score of those students in the control group who used the traditional language laboratory.

Table 4. Comparison of mean scores of CALL (n=100) and control (n=172) groups on grammar test items on final exam.

	CALL	Control	F	level of significance
mean score	82.2	73.4	49.4	.0001

Table 5. Comparison of mean scores of CALL (n=57) and control (n=119) groups on vocabulary test items on final exam.

	CALL	Control	F	level of significance
mean score	79.7	70.0	16.4	.0001

For the grammar section, the mean score in each achievement group (low, medium, and high), was also higher for the experimental group than for the control group, with significant differences for the "middle group" and the "high group" (Table 6).

Table 6. Comparison of mean scores of CALL and control groups on grammar test items on final exam for low (n=42), middle (n=134), and high (n=96) subgroups.

Subgroup	CALL	n	Control	n	t	level of significance
low	66.6	14	60.6	28	1.67	.10
middle	80.7	50	71.4	84	4.07	.0001
high	90.3	36	82.2	60	3.34	.001

Similarly, the mean score in each achievement group (low, medium, and high) was also higher for the experimental group than for the control group on the vocabulary test items, with significant differences for the "low group" and for the "middle group" (Table 7).

Table 7. Comparison of mean scores of CALL and control groups on vocabulary test items on final exam for low (n=33), middle (n=90), and high (n=53) subgroups.

Subgroup	CALL	n	Control	n	t	level of significance
low	76.3	8	57.2	25	2.35	.026
middle	75.5	28	57.2	62	2.21	.029
high	86.5	21	84.8	32	0.53	.596

On the separate vocabulary test, the overall mean score was higher for those words taught through CALL than through traditional methods (Table 8).

Table 8. Comparison of mean scores of computer group on words taught through CALL (number of words=57) and traditional methods (number of words=57) on separate vocabulary test.

	words via CALL	words via traditional	F	level of significance
mean score	83.6	74.7	5.48	.0054

Mean scores were also significantly higher for all three achievement groups in the experimental group for words taught by CALL than for words taught through traditional methods (Table 9). According to Avent, "in this study it is clear that when the students learned vocabulary items by computerized instruction that, without exception, they remembered them better than the words which had been learned using traditional methods" (pp. 82-83).

Table 9. Comparison of mean scores of low, middle, and high subgroups of experimental group for words taught through CALL and words taught through traditional methods.

Subgroup	words via CALL	words via traditional	t	level of significance
low	84.1	68.3	5.20	.0013
middle	80.9	71.4	3.34	.0023
high	88.9	82.8	2.26	.037

As stated earlier, students in the experimental group on average spent two hours more in the computer lab than students in the control group spent in the language lab. These disparate times may help explain the differences in achievement scores at the end of the study between students in the two groups, and poses an intriguing question. On the one hand, students using CALL reached higher levels of achievement in German through extended practice in the computer

lab. On the other hand, CALL required more time than traditional instruction in this study. Avent himself addresses this issue in his conclusion:

> The information provided by this study does, it seems, indicate that computer-assisted language learning is effective. It works. Whether or not it is efficient is still somewhat open to question. Regardless of the efficiency or lack thereof, if the goal is for the student to learn the material, then the result of this study would indicate that computer-assisted language learning is a viable alternative, and its development should be pursued. (pp. 96-97)

In another study, Nieves converted *EXITO*, a multimedia CALL program in Spanish originally developed by the Central Intelligence Agency, into a one-semester college course in introductory Spanish and conducted a formative evaluation of this course under development. Nieves inverted the typical ratio of computer time to classroom time in this experimental course, as students spent four to five hours per week in the computer lab and only one hour per week in class with the instructor. In addition to the largely qualitative formative evaluation of *EXITO*, Nieves included a modest pilot study in this 1994 dissertation to compare the performance of 19 students in the CALL group with another 18 students in a control group on a Spanish proficiency exam. In comparing the mean scores (out of a possible 160 points) of the students in the CALL and control groups, Nieves found that students in the CALL group scored somewhat higher (CALL mean=97, control mean=90) and had a much smaller range between the highest and lowest scores (CALL range=65, control range=112). Further, when Nieves broke down these mean scores by "true beginners" and "false beginners," with the former representing students who had never studied Spanish before, the difference between the *EXITO* and control groups became more pronounced. The mean score on the proficiency exam was substantially higher for "true

beginners" in the CALL group (mean=84) than for "true beginners" in the control group (mean=60).

Wright's study, a 1992 master's thesis, compared student achievement on three chapter tests in beginning German with an experimental group of 45 students using a computerized workbook and a control group of 62 students using a standard workbook for vocabulary and grammar study. Computerized workbooks and standard workbooks provided similar content and exercises, with computerized workbooks also able to give instant feedback and suggestions for finding correct answers. These computerized workbooks could also help explain to the student why an answer was correct and listed the page number in the textbook where an explanation could be found. By contrast, the standard workbook provided only an answer key to the questions at the back of the book without explanation. Students in both experimental and control groups still used standard workbooks for listening and communication exercises. The mean scores were higher on all three chapter exams for the CALL group (Table 10).

Table 10. Comparison of mean scores of CALL and control groups on three chapter exams.

	CALL	n	Control	n
Chapter 1	84.9	49	79.6	59
Chapter 2	85.3	48	80.4	62
Chapter 3	86.4	49	85.0	42

Wright's findings need to be approached with caution, however. Nonequivalence between experimental and control groups in this study make significance tests inappropriate. Assignment to the experimental and control groups was performed on the level of section, whereas assignment

on the level of the individual would have allowed better comparisons between the two groups. In addition, only one experimental section was randomly chosen from a group of seven sections, and the other two experimental sections were taught by Wright himself. Since Wright was involved on such a personal level with the pilot project, he acknowledged that "it was impossible to eliminate teacher/researcher bias completely" (p. 55). The superior performance of students in the experimental group could be plausibly attributed to the personal involvement of Wright as teacher rather than to the efficacy of the CALL workbook program.

Stenson et al.'s 1992 CALL study analyzed the progress in overall English pronunciation--including stress, rhythm, and intonation--of two groups of international teaching assistants enrolled in a quarter-long course to improve their spoken English. One group of 18 students used SpeechViewer, an IBM software program which provides visual representations of speech, as part of the class, and a control group of 35 students worked with more traditional methods of pronunciation practice. Stenson and her colleagues make no mention of randomization in the assignment of individuals to experimental or control groups. Students in the CALL and control groups attended one two-hour group session each week, with four students assigned to each group session. Each student also received 50 minutes of one-on-one instruction every week. Students in the CALL group had instructors who regularly used SpeechViewer in the one-on-one tutorials, while students in the control group did not use CALL at all during their 50-minute sessions. For students in the CALL group, the average session on SpeechViewer lasted 15 minutes during a 50-minute tutorial session, and the average total amount of time with SpeechViewer for these students over the quarter was 80 minutes.

Stenson et al. assessed student pronunciation performance using an exam called SPEAK, commercially available through the Educational Testing Service, and the "Mimic Test," a test of English language designed for the study by the researchers in which students were asked to listen to a native speaker pronounce words, phrases, and sentences, and then repeat them, mimicking the model as closely as possible. Despite claims of general widespread enthusiasm for SpeechViewer, no substantial differences were found between pre- and post-test scores for the CALL and control groups on both the SPEAK and Mimic tests. Stenson speculates that the international teaching assistants in the CALL group "simply did not get enough practice with SpeechViewer to show dramatic results" (p. 13).

This last finding based on the empirical CALL studies provides a link to the CALL reviews that also constitute part of this restrospective analysis. Prior to 1990, comparisons of students using CALL with students using traditional methods of language learning appeared more frequently than in recent years. At the time, researchers apparently felt more concern about establishing the efficacy of CALL. The overview of reviews that follows serves as a foundation for the previous analysis of empirical CALL studies by providing a summary of the state of CALL research up to 1990, offering a benchmark against which to assess the direction of current CALL inquiry.

Part Three: Review of CALL and CAI Reviews

State of CALL Research up to the Early 1990s

Up to the early 1990s, CALL research had yielded no consistent, unambiguous, and definitive findings. Several reviewers concluded that too few studies without obvious validity

problems were available for close examination (Niemiec & Walberg, 1987; Roblyer et al., 1988; Dunkel, 1991; Chapelle & Jamieson, 1990; Garrett, 1991). Pederson (1987) placed the state of the research enterprise in CALL in the late 1980s into an historical context by comparing it with the experience of language teachers in the early 1960s when the language lab was the emerging technology in language teaching. She asserted that the language lab failed to live up to its high expectations in large part due to the lack of good research on how to use the technology for language learning. In the late 1980s, CALL software designers were similarly handicapped by an inadequate research base.

Smith (1987) offered several observations as explanations for this state of affairs in the CALL research enterprise. First, CALL appeared during the backlash against the behaviorist theoretical underpinnings of the language lab, and, as a result, through the 1980s many second language teachers viewed CALL with a good deal of skepticism. The second language acquisition (SLA) theory current in 1987 rested not on behaviorist theories, but instead viewed language learning as the development of a functional communication ability in the target language rather than as simple acquisition of a vocabulary and the rules of grammar. For example, following the tenets of behavioral theory, the Audio Lingual Method, most common to the language lab, trained students to utter correct sentences in the second language (L2) by memorizing and repeating L2 dialogues from a series of audiotapes. By contrast, the functional communication aspect of SLA theory stresses that one avenue of student language acquisition occurs as a consequence of spontaneous conversation in the L2, the purpose being to communicate with a partner in a conversation. This "functional communication" can occur without formally correct grammar or vocabulary as the partners both modify their L2 utterances

in negotiating meaning in the target language. Most CALL programs were not designed to use this "functional communication" paradigm.

Second, because few L2 teachers in the mid-1980s used CALL effectively, few could therefore serve as models or mentors to others. Even if more good examples of the application of CALL had been available, Smith asserted that few teachers in the L2 teaching force of the late 1980s were disposed toward personal computer literacy and toward pedagogical computer literacy using CALL. Furthermore, asserted Smith, many language teachers who did use CALL were not trained to use it effectively.

Third, many CALL programs themselves were flawed. Many L2 teachers who created their own CALL programs generally lacked the technically sophisticated computer programming skills necessary to produce CALL lessons that students would regard as high in quality when compared with other programs in the students' experience (Smith, 1987). Conversely, computer program specialists who worked in CALL generally lacked a deep understanding of the theory and pedagogy of SLA. In addition, we found no reports of L2 teachers or CALL programmers working closely with instructional design specialists in the creation of CALL programs.

Several reviews suggested that because CALL had not yet become a mature research field, the validity of a number of primary CALL studies was questionable because of problems in research design and execution (Chapelle and Jamieson, 1990; Pederson, 1987; Williams and Brown, 1991).

The most common objection noted in the reviews in both CAI and CALL research was to the simple research design of computer vs. non-computer. Many studies thus reported that the differences in experimental outcomes were due to the computer per se rather than to specific

features of the CAI or CALL lesson, or possibly to characteristics of the students, or to the nature of the subject matter, or to interactions among these variables (Williams & Brown, 1991; Pederson, 1987).

In spite of the relative immaturity of research on computer-assisted instruction (CAI), investigators could identify some trends by the late 1980s. The most important of these was that CAI seemed to work: students who used CAI in various subjects areas achieved more than students who experienced only traditional classroom instruction.. An average effect size for CAI of about .36 derived from two extensive CAI meta-analyses (Kulik & Kulik, 1991; Niemiec & Walberg, 1987) indicated that the median student scoring at the 50th percentile in a traditional classroom would score at the 64th percentile, on average, if he or she used computer-assisted instruction. (See box for more thorough explanation of effect size.)

EFFECT SIZES

A common method of reporting results of empirical studies and meta-analyses is to use effect sizes. In brief, an effect size is a simple way to compare the outcomes among studies with differing numbers of participating students by standardizing the results. For example, an effect size of .5 would mean that, on average, students formerly at the 50th percentile would now achieve at the 69th percentile. An effect size of 1.0 would move the median student to the 84th percentile. Educators generally agree that an effect size of around .3 (a move to the 62nd percentile for the median student) or larger represents a substantial education benefit, especially when we consider that these effect sizes represent the average improvement for a population of students, not just one student. However, the merit of a given effect size for any education intervention also depends on what other options may be available and on their relative costs.

SAMPLE EFFECT SIZES AND THEIR RELATED PERCENTILE DIFFERENCES

EFFECT SIZE	.00	.10	.20	.30	.40	.50	.75	1.0	1.5	2.0
PERCENTILE	50	54	58	62	66	69	77	84	93	98

Not all researchers in the field expressed enthusiasm about the computer vs. non-computer research designs that produced the average effect sizes reported above, suggesting that it was not the computer itself that is the cause of better student achievement, but the way that the computer

BEST COPY AVAILABLE

lessons were structured that led to increases in student achievement (Dunkel, 1991; Pederson, 1987; Williams & Brown, 1991).

Researchers and reviewers who question the simple computer vs. non-computer design are, in our judgment, thoughtfully suggesting that the computer is not a magic bullet. Rather, computer lessons contain a number of variables, and each variable needs to be identified, operationalized, and examined in order to determine what it is about CALL lessons that result in improved student achievement and attitudes. Furthermore, different lesson variables affect different students in different ways.

Other Comments on CAI and CALL Research

Williams and Brown (1991) noted that another common problem in CAI research is that many studies did not explicitly base their research design on any particular theory of learning. As a result, the researcher lacked justification for attributing the outcome to any particular aspect of the experimental treatment (Pederson, 1987; Chapelle & Jamieson, 1989).

In addition, Williams and Brown expressed disappointment in much of the reporting of CAI studies (1991). Many studies did not describe fully the experimental treatment, the students, or the CALL lessons. Pederson added that many researchers failed to define adequately and to operationalize the variables the study purported to examine (1987). It is difficult, therefore, to judge the results of such studies.

Williams and Brown (1991) called attention to a subtle interaction among the components of the computer medium, the individual characteristics of the learner, and the specified learning outcomes of the lesson. This interaction precludes treating the instructional medium, be it teacher-lecture, interactive-video, or a computer, as a cohesive whole, and therefore treating

experimental outcomes as if they were caused by "the lecture," or by "the video," or by "the computer." In other words, different aspects of lessons on the computer for different subject areas have different effects on different students.

Two reviews provided guidelines for addressing this issue of the quality of research reporting in CAI and CALL. For CAI reporting, Roblyer (1988) suggested that, at a minimum, the researcher provide the reader with an adequate description of the experimental design (including sample sizes of the full sample and any sub-groups), information on any testing instrumentation used, more complete statistical data than is sometimes reported, and a description of the experimental treatment so that the reader can understand what was done.

Chapelle and Jamieson (1989) added these suggestions. First, CALL researchers should provide the reader with the SLA theory that informed the study, explain how it applies to the learners in the study, and include a description of the kinds of cognitive processing that this CALL lesson is intended to stimulate. Second, researchers should provide a description of the learners, for example, their prior language learning experience and demographic information (age, gender, ethnicity, grade in schooling). Third, researchers should provide the reader with a description of the CALL lessons used, including at least the following information: type of activity (e.g., drill, game, simulation), planned learning outcome(s), learner focus (what the student is actually doing), the linguistic purpose of this lesson, level of the lesson (e.g., beginner, novice, intermediate, advanced), the lesson's degree of tolerance for different levels of performance by different students, and a description of how the teacher integrated the lesson into the course. Fourth, when reporting the outcomes of the lesson, researchers should include a description of the learning strategies that the students appeared to use in response to the lesson.

Attitude

Two of the thirteen reviews reported that student attitudes toward computer learning in all subjects are more positive than toward traditional classroom lessons [average effect sizes = .62 (Roblyer et al., 1988) and .28 (Williams & Brown, 1991)]. These two reviews also reported that students' attitudes toward school were more positive when CAI was part of their school experience (average effect sizes, .22 and .33, respectively). Roblyer et al. (1988) caution, however, that in their opinion, the positive student attitudes reported may not justify the cost of the hardware, software, and teacher training to establish CAI as an integral part of students' school experience.

Design of CAI and CALL Experiences for Students

The most consistent finding reported in the research on CAI and CALL up to the early 1990s is that although CAI and CALL are effective in improving student achievement when used as a supplement to traditional classroom instruction, neither is apparently effective as a *replacement* for traditional classroom instruction (Roblyer et al., 1988; Robinson, 1991; Williams & Brown, 1991).

Language learning is a socially mediated activity, and CALL introduces the computer and other technology as another "player" in the constellation of social interactions in a second language (L2) classroom (Johnson, 1990). Johnson reported that the group size that promotes the most effective social interactions in the target language when a computer is a member of the group is 2 or 3 students, especially if the CALL program assigns specific roles to each of the students (citing reports from three investigators). A promising example is collaborative composition among 2 or 3 students using a word-processing program in the target language.

Use of the computer in the form of local area networks (computers networked within an L2 classroom) and wide area networks (computer access to the Internet), shows promise for extending the kinds of social interactions possible in the target language (Robinson, 1991; Scott, et al., 1992). For example, students in an American classroom can converse through e-mail with students in a classroom halfway around the world in the language of those other students.

The specific way CALL lessons are put together, what CALL users and programmers refer to as the program's *coding elements*, are important variables in the efficacy of CALL lessons. For example, CALL programs that require students to make extended responses (rather than just type the "enter" key) result in higher student achievement (Chapelle & Jamieson, 1990). Similarly, CALL programs that require more interaction between the student and a video with native speakers result in higher student achievement.

Traditional drill and practice CALL programs are more effective if students must understand the meaning of the L2 sentences in which the grammatical corrections are to be made, instead of simply making the corrections mechanically in sentences whose meaning they need not understand in order to complete the exercise (Pederson, 1987; Chapelle & Jamieson, 1991).

CALL software that leads students to think in new ways, that is, that gives them new patterns of cognition in relation to the target language, are more effective (Pederson, 1987). For example, Robinson (1991) asserts that theory suggests that CALL programs that support implicit error correction by students, that is, leading students to find and correct their own mistakes rather than having mistakes highlighted or otherwise flagged by the computer, will result in higher student achievement.

Finally, programs that lead students to share in the control of the lessons result in higher student achievement. A midway approach with regard to student control of help menus, that is, help functions of the CALL program that are neither totally controlled by the CALL program itself, nor completely in the hands of the student, results in better student achievement in the target language (Robinson, 1991). For example, an error message that appears on the screen after a student has typed in a sentence in the target language that says simply, "agreement?" could lead the student to examine the sentence for errors in noun-adjective agreement, subject-verb agreement, or noun-pronoun agreement. A student unable to discover the agreement mistake would need to ask for further assistance.

Other Outcomes of CAI and CALL Lessons

Other outcome variables beyond student achievement and student attitude that have been studied include student learning time, course completion rates, retention time, and cost factors for using CALL (Williams & Brown, 1991; and Niemiec & Walberg, 1987).

Scott et al. (1992) reported that the results of the Apple Classroom of Tomorrow study suggest that CAI produces more positive student interactions, such as spontaneous peer tutoring and cooperative learning, and in addition, leads students to become more active learners . Two evaluations of the Apple Classroom of Tomorrow study report that students took more initiative and assumed more responsibility for their own learning when using CAI in the Apple Classroom of Tomorrow.

Teachers and CAI/CALL

Effective integration of CAI or CALL lessons into a curriculum requires that teachers learn a new role, that they learn a new pedagogy that differs from their former teaching

methodology. Computers in a classroom can put into the hands of the students more control of their own learning, shifting that control away from the teacher. Scott et al. (1992), reporting again on the Apple Classroom of Tomorrow study, note that new teacher behaviors appeared only after the teachers had solved the new management problems presented by the computer-rich classroom environment. For example, teachers need to discover new ways to keep track of students' learning when each student may be doing different work. Therefore, they conclude, teachers will need much support during the introduction of CAI and CALL lessons into any curriculum.

Pederson (1987) supports this conclusion, noting that the results of teacher surveys on CALL indicate that L2 teachers strongly desire additional and better training in the use of CALL in their classrooms. The same surveys indicate that teachers were not at all satisfied with the CALL software available in the mid-1980s.

Conclusions From the Reviews

In sum, CALL reviewers before the early 1990s had consistently made a handful of recommendations for future CALL researchers. Overwhelmingly, they called for researchers to abandon the simple CALL vs. non-CALL research design and to focus more specifically on finding what components of CALL lessons are effective with what kind of language lessons for what kind of students. That is, what is the nature of the interactions among student characteristics, CALL lesson design, desired learner outcomes, and computer coding capabilities? Two substantial subsets of this general recommendation are specific recommendations (1) to look at components of CALL lesson design (e.g., program branching, error analysis and feedback, screen design) and (2) to examine student characteristics (e.g., learning style, cognitive approach, gender) as they relate to CALL effectiveness.

A number of reviewers suggested that the impact of CALL research would be much enhanced if investigators explicitly designed their studies around theories of linguistics and/or cognitive development. Several reviewers noted that the power of the computer could be harnessed as a research tool in "observing" student learning behavior by recording all keystrokes made by students during a CALL lesson. Finally, several suggested that researchers examine the cost-effectiveness of CALL, especially in comparison with other kinds of language teaching.

Many of the researchers whose studies we reported in Part Two of this paper have answered these calls of the reviewers. (See Table 11) Twenty one of the twenty-two studies have in some way examined components of CALL lesson design, linguistic outcomes, or student characteristics, and the interactions among these components. Twenty examined CALL lesson design charateristics in a specific way. Twenty either stated a particular theoretical basis for their study design, or reflected a

Table 11. Number of studies (n=22) that addressed research issues suggested in the 13 reviews.

SUGGESTED RESEARCH ISSUES TO EXAMINE	FREQUENCY
Interactions among student, lesson, computer, context	21
Lesson design, computer coding elements	20
Theory	20
Student characteristics	11
Comparison with other instructional interventions	10
Record of student behavior during lessons, e.g., keystrokes	9
Cost compared to other instructional strategies	2

particular theoretical stance in their design, even though it may not have been specifically identified in the report of the study. Eleven closely examined student characteristics and their influence on CALL outcomes. Ten compared CALL efficacy with other instructional interventions not mediated by a computer. Nine reported using the computer's capabilities to record student keystrokes as the researchers gathered data to explore aspects of various cognitive or linguistic theories.

Only two studies addressed the issue of cost, and these did so only in passing. Both studies examined the interactions of CALL and cooperative learning; both found no difference in student achievement between individual student computer use and pairs of students in a cooperative learning situation at the computer. Both suggest that since pairs of students seem to learn no less than do individual students, schools for whom budget constraints are an issue may safely consider purchasing half as many computers for their language classes that use CALL.

As noted earlier, several reviewers noted the questionable validity of many CAI and early CALL studies. These reviewers made a number of suggestions to improve the validity of future studies. We emphasize some of those suggestions here. In doing so, we note the complexity of improving internal and external validity in so complex an endeavor as CALL.

In designing a study, researchers should account for as many variables as possible that may impact student performance on the L2 measure that is being examined. Variables should be carefully defined and operationalized. In addition, variables from all three streams should be included in the study design: desired linguistic learning outcomes (informed by theories of language learning), instructional design of the lesson (informed by theories of cognitive psychology), and the computer coding elements used in the lesson. Finally, variables accounting for student characteristics should be included in the study design.

The twenty-two empirical studies we examine here exhibited varying degrees of success in addressing these validity issues. We note in particular two studies that did so especially well: Avent's study comparing achievement for students using either the traditional language lab or CALL lessons, and Garza's study of the effect of captioning segments of "authentic" video on students' recall of the dialogue in the video segments. (See pages 13-14 for a detailed description of Garza's study, and pages 27-30 for Avent's study.)

Part Four: Recommendations for the Future of CALL

Based on this retrospective analysis of CALL research, we present here three general conclusions, each accompanied by recommendations for future CALL practice and research:

1. Good CALL programs are hard to find because integrating the three streams of educational psychology, linguistics, and computer technology is difficult to do well.

Relatively few individuals have expertise in all three areas of educational psychology, linguistics, and computer technology. Since effective CALL programs seem to require successfully integrating these three streams, we recommend that CALL developers consider working in creative teams and combining different types of expertise when authoring and implementing CALL programs. An example of this kind of model in practice in another technology-conscious field is the children's television show *Sesame Street*, which from its inception developed programming by bringing together a collaboration of television writers and

producers, classroom teachers, professors, researchers with expertise in evaluation, songwriters, and animators (Lesser, 1974).

We also recommend that educators in institutions of higher education review CALL programs that have already been developed -- especially in federally-funded organizations -- and investigate the possibility of converting these programs for classroom use. Foreign-language programs in agencies such as the Central Intelligence Agency (CIA), the Defense Department, the Foreign Service, and the Peace Corps generally represent considerable investments of time and money, and some of these programs may contain CALL or other technological components appropriate for domestic spinoffs in colleges and universities. The Spanish-language program *EXITO*, originally developed by the CIA, is an example of this conversion of CALL from federal agency to college classroom.

Representing an investment of millions of dollars and several years of development, *EXITO* was created by the CIA's Foreign Language Training Laboratory in 1985 as an intensive 10-day course to teach survival Spanish to CIA agents at a proficiency level equivalent to about a year of college Spanish (*Washington Post*, April 3, 1994). A multimedia CALL program, *EXITO* features native speakers in vignettes and lessons that integrate video, audio, graphics, and animation. A workbook and set of six audiotapes supplement the *EXITO* software. For each of the ten days in the program, participants are expected, at minimum, to study four hours on the computer, spend one hour in a one-on-one tutorial with an instructor, perform written exercises for one hour in the accompanying workbook, and work one hour with the audiotapes (Nieves, 1994). *EXITO* uses laserdisk technology, and the computer workstations required to support

EXITO in its original form each cost several thousand dollars (*Computer Reseller News*, September 27, 1993).

EXITO, however, is now available to universities and schools in a format that requires only a personal computer equipped with a double-speed CD-ROM plus a Motion Picture Experts Group (MPEG) video board, which sold in 1995 for about $400. Several years ago, the CIA entered into an agreement with Analysas, a private, Washington, DC-based company, to develop *EXITO* for educational and commercial use. According to the agreement, the CIA receives royalties for sales to private organizations, but does not profit from sales to schools, which can purchase the *EXITO* package at discount rates. Kelly Ann Nieves' 1994 dissertation, *The Development of a Technology-Based Class in Beginning Spanish: Experiences With Using* EXITO, chronicles the process of converting this recent version of *EXITO* into a one-semester Spanish course at George Mason University (Nieves, 1994).

2. New multimedia computer technologies offer ways to develop more CALL programs that emphasize watching, listening, and speaking in addition to the traditional CALL activities of reading and typing, and new networking computer technologies provide opportunities to use CALL to promote person-to-person interaction in the target language that transcend traditional obstacles of distance and time.

Whereas earlier versions of CALL basically consisted of a closed loop between student and machine and emphasized textbook-style reading and writing activities, educators in colleges and universities can now use CALL both as a vehicle for engaging students in watching, listening,

and speaking activities and for connecting students with other people outside of the classroom for conversations in the target language.

Captioning video segments in the target language represents one way of leveraging new multimedia computer technologies into improved student foreign-language learning, where the simultaneous presentation of language in spoken and written forms provides more comprehensible input in the target language by engaging both a student's aural and visual sensory receptors. In general, multimedia CALL provides opportunities for students to learn languages in more authentic contexts, as students can observe native speakers in ordinary situations in foreign countries and interact with the CALL program in a variety of manners.

The use of CALL as a vehicle for interpersonal communication in the target language over computer systems such as the Internet allows individual, subjective perspectives on topics of mutual interest to surface. Since participants in these conversations around shared interests usually focus on the content, or meaning, of language rather than its form, this application of CALL corresponds with the humanistic and cognitive currents in the stream of educational psychology. This interaction in the target language via CALL is also consistent with the Natural Approach, a current in the linguistics stream, which posits that second language learning occurs more effectively when people feel more invested in what they want to say than anxious over how correctly they say it.

3. <u>The field of CALL needs more research, especially formative evaluation, conducted by a larger pool of researchers</u>.

In the course of conducting our review on CALL in higher education in the U.S., we noted that much of the research has been capably conducted by two groups: (1) a relatively small group of researchers whose names appear repeatedly in the literature, and (2) a group of graduate students writing their theses on the subject. Given the potential importance of CALL in colleges or universities, we wonder if the responsibility for CALL research should continue to fall on so few shoulders.

Furthermore, more coordination of CALL research around a better-defined agenda seems highly desirable. If practitioners and researchers in the field of CALL could agree upon a set of key questions for subsequent studies to address, the resulting literature might provide stronger guidance concerning how to use computers to improve teaching and learning in foreign languages in colleges and universities.

Finally, educators might consider the allocation of more resources for formative evaluation in order to investigate the effectiveness of specific CALL programs with particular students at particular sites. Because the "CALL" designation covers a wide variety of programs that can be very dissimilar from one another in their standpoints in relation to educational psychology, linguistics, and computer technology, foreign language educators may find it helpful to complement the insights offered in the general CALL literature with formative research on programs-in-development.

References: General, Empirical Studies, and Reviews

General

The Caption Center. (1995). *Tech Facts*. Vol. 4. Boston: WGBH Educational Foundation.

Computer Reseller News. Forming a more perfect union. September 27, 1993.

Digest of Educational Statistics. (1993). Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Ely, Donald P. (1993). Computers in schools and universities in the United States of America. *Educational Technology*, 33 (9), pp. 53-57.

Fox, Jeremy. (1991). Learning languages with computers: A history of computer assisted language learning from 1960 to 1990 in relation to education, linguistics, and applied linguistics. Doctoral Thesis, University of East Anglia (England).

Lesser, Gerald. (1974). *Children and television: Lessons from Sesame Street*. New York: Random House.

Munson, Janet R., Richter, Randy L., and Michael Zastrocky. (1994). *CAUSE Institution Database: 1994 Profile*. Boulder, CO: CAUSE.

Statistical Abstract of the United States. (1994). Washington, D.C.: Government Publications Office.

Washington Post. Speak Spanish like a spy. April 3, 1994.

Empirical CALL Studies Published Since 1989

Aspillaga, Macarena. (1991). Screen design: Location of information and its effects on learning. *Journal of Computer-Based Instruction*. 18 (3), pp. 89-92.

Avent, Joseph. (1993). *A Study of Language Learning Achievement Differences Between Students Using the Traditional Language Laboratory and Students Using Computer-Assisted Language Learning Courseware*. Doctoral Thesis, University of Georgia.

Bationo, Bernadin. (1992). The effects of three feedback forms on learning through a computer-based tutorial. *CALICO Journal*. 10 (1), pp. 45-52.

Borras, Isabel & Robert Lafayette. (1994). Effects of multi-media courseware subtitling on the speaking performance of college students of French. *The Modern Language Journal*, 78 (1), pp. 61-75.

Borras, Isabel. (1993). Developing and assessing "Practicing Spoken French." *Educational Technology Research and Development*, 41 (4), pp. 91-103.

Chang, Kuan-Yi & Smith, Wm. Flint. (1991). Cooperative learning and CALL/IVD in beginning Spanish: An experiment. *The Modern Language Journal*, 75 (2), pp. 205-211.

Chapelle, Carol & Suesue Mizuno. (1989). Students' strategies with learner-controlled CALL. *CALICO Journal*, December issue, pp. 25-47.

Chun, Dorothy M. (1994). Using computer networking to facilitate the acquisition of interactive competence. *System*, 22 (1), pp. 17-31.

Fischer, Robert. (1989). Instructional computing in French: the student view. *Foreign Language Annals*, 26 (4), pp. 527-534.

Garza, Thomas. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals*, 24 (3), pp. 239-257.

Hermann, Francoise. (1992). *Instrumental and Agentive Uses of the Computer: Their Role in Learning French as a Foreign Language*. San Francisco: Mellen Research University Press.

Hsu, Jing-Fong, Chapelle, Carol & Ann Thompson. (1993). Exploratory learning environments: what are they and what do students explore? *Journal of Educational Computing Research*, 9 (1), pp. 1-15.

Jameison, Joan, Campbell, Joan, Norfleet, Leslie & Berbisada, Nora. (1993). Reliability of a computerized scoring routine for an open-ended task. *System*, 21(3), pp. 305-322.

Jamieson, Joan, Norfleet, Leslie & Berbisada, Nora. (1993). *Successes, failures, and dropouts in computer-assisted language learning*. ERIC ED 354 786.

Mitchell, Cristi. (1992). *The Relationship of Computer-Assisted Language Learning Environments and Cognitive Style to Achievement in English as a Second Language.* Doctoral Thesis, University of Miami.

Nagata, Noriko. (1993). Intelligent computer feedback for second language instruction. *The Modern Language Journal*, 77 (3), pp. 330-339.

Nieves, Kelly. (1994). *The Development of a Technology-Based Class in Beginning Spanish: Experiences with using* EXITO. Doctoral Thesis, George Mason University.

Raschio, Richard. (1990). The role of cognitive style in improving computer/assisted language learning. *Hispania*, 73 (May), pp. 535-541.

Shiu, Ka-Fai & Sharon Smaldino. (1993). *A pilot study: Comparing the use of computer-based instruction materials and audio-tape materials in practicing Chinese.* ERIC ED 362 204.

Stenson, Nancy, Downing, Bruce, Smith, Jan & Smith, Karin. (1992). The effectiveness of computer-assisted pronunciation training. *CALICO Journal*, Summer issue, pp. 5-19.

Wright, David Allan. (1992). *The reciprocal nature of universal grammar and language learning strategies in computer assisted language learning*. Masters Thesis, University of Arizona.

Reviews of CALL and CAI

Chapelle, Carol, and Jamieson, Joan. (1991). Internal and external validity issues in research on CALL effectiveness. In *Computer-Assisted Language Learning and Testing: Research Issues and Practice.* Dunkel, Patricia, editor. New York: Newbury House. pp. 37-60.

Chapelle, Carol, and Jamieson, Joan. (1989). Research trends in computer-assisted language learning. In *Teaching Languages With Computers: The State of the Art*. Pennington, Martha, editor. La Jolla, CA: Athelstan. pp. 45-60.

Dunkel, Patricia. (1991). The effectiveness research on computer-assisted instruction and computer-assisted language learning. In *Computer-Assisted Language Learning and Testing: Research Issues and Practice.* Dunkel, Patricia, editor. New York: Newbury House. pp. 5-36.

Garrett, Nina. (1991). Technology in the service of learning: Trends and issues. *Modern Language Journal*, 75, (1), pp. 74-96.

Johnson, Donna M. (1991). Second language and content learning with computers: Research in the role of social factors. In *Computer-Assisted Language Learning and Testing: Research Issues and Practice.* Dunkel, Patricia, editor. New York: Newbury House. pp. 61-84.

Niemiec, Richard, and Walberg, Herbert J. (1987). Comparative effects of computer-assisted instruction: a synthesis of reviews. *Journal of Educational Computing Research*, 3, (1), pp. 19-37.

Pederson, Kathleen M. (1987). Research on CALL. In *Modern Media in Foreign Language Education: Theory and Implementation*. Smith, Wm. Flint, editor. Lincolnwood, IL: National Textbook Company. pp. 99-131.

Robinson, Gail L. (1991). Effective feedback strategies in CALL: learning theory and empirical research. In *Computer-Assisted Language Learning and Testing: Research Issues and Practice*. Dunkel, Patricia, editor. New York: Newbury House. pp. 155-166.

Roblyer, M.D., Castine, W.H., and King, F.J. (1988). *Assessing the Impact of Computer-Based Instruction*. New York: The Haworth Press.

Scott, Tony, Cole, Michael, and Engle, Martin. (1992). Computers and Education: A Cultural Constructivist Perspective. In *Review of Research in Education - 18*. Grant, Gerald, editor. Washington, DC: American Educational Research Association. pp. 191-254.

Smith, Wm. Flint. (1988). Modern media in foreign language education: A synopsis. In *Modern Media in Foreign Language Education: Theory and Implementation*. Smith, Wm. Flint, editor. Lincolnwood, IL: National Textbook Company. pp. 1-12.

Roblyer, M.D., Castine, W.H., and King, F.J. (1988). *Assessing the Impact of Computer-Based Instruction*. New York: The Haworth Press.

Williams, Carol J., and Brown, Scott W. (1991). A review of the research for use of computer-related technologies for instruction: An agenda for research. In *Educational Media and*

Technology Yearbook - 1991. Branyan-Broadbent, Brenda and Wood, R. Kent, editors. Englewood, CO: Libraries Unlimited. pp. 26-46.

Appendix A
Search Procedure

We searched for pertinent material on computer-assisted language learning (CALL) primarily through four databases: the Educational Resources Information Center (ERIC), the Harvard on-line card library card catalog system (HOLLIS), the Social Sciences Citation Index (SSCI), and Dissertation Abstracts International (DAI).

We used ERIC both on-line (1989-present) and on CD-ROM (1986-1994) and identified articles and papers potentially relevant to CALL employing keyword and subject category searches related to computer-assisted instruction, foreign language instruction, evaluation research, and educational technology. One specific search, which retrieved 72 items, was [(FOREIGN and LANGUAGE and TECHNOLOGY and EVALUATION and INSTRUCTION not (LANGUAGE LABORATORIES and EVALUATION)]. Other ERIC searches included KW COMPUTER ASSISTED LANGUAGE INSTRUCTION (n=21) and SU TECHNOLOGY (n=870). We then inspected the abstracts for these materials and retrieved the hardcopy (either microfiche or journal articles) for those papers that either (1) provided useful historical or theoretical background information on CALL or CAI or (2) offered original studies or reviews of studies that evaluated the use of CALL in various language-learning situations.

We used HOLLIS to identify books and chapters dealing with CALL. Searches included [KW FOREIGN LANGUAGE and KW COMPUTER] (n=11), [KW COMPUTER ASSISTED and KW FOREIGN] (n=3), SU LANGUAGE AND LANGUAGES--COMPUTER ASSISTED INSTRUCTION (n=27). We then retrieved these books from Gutman and Widener libraries, and searched the bibliographies of these books for materials about CALL that did not appear in our ERIC

and HOLLIS searches. We obtained the hardcopy for those materials we found to appear related to CALL.

Throughout this initial search for CALL materials, we focused on materials published in the 1980s and 1990s. Because computers are a relatively new technology in education, little literature on the topic exists before 1980. We discovered that even literature published in the early 1980s about CALL was of limited relevance to current uses of computers in the foreign language classroom due to dramatic differences in hardware (and to some extent software) during the subsequent ten years.

After reviewing the materials that emerged from our initial ERIC and HOLLIS searches of the CALL literature, we decided to narrow the focus of our inquiry. Specifically, we selected six criteria for inclusion in our review:

(1) original studies of CALL

(2) in college and university settings

(3) in the United States

(4) published since 1989

(5) that considered the differential achievement of at least one group of students

(6) by analyzing at least one quantitative measure of student performance.

We chose 1989 as our cutoff date for original studies for three main reasons. First, we felt that technological advances in CALL hardware and software in recent years raised questions about the relevance of these earlier studies to the situation of CALL in the mid-1990s. Second, we found few studies published before 1989 that satisfied our first three criteria. Third, we discovered that several authors had already written comprehensive summaries of CALL research up to 1989 (e.g.,, Dunkel,

1991). The decision to focus on CALL in higher education in the United States reflected our own substantive interests.

In searching for additional CALL studies that met our six criteria and had not surfaced in our initial ERIC and HOLLIS searches, we turned to two more reference databases, the Social Sciences Citation Index (SSCI) and Dissertation Abstracts International (DAI). The SSCI, available to us both in print and on CD-ROM, allows one to trace studies forward by listing every article (among the more than 1000 journals in the database) that cites a particular work in its list of references. We used SSCI to generate lists of the articles published in the social science literature between 1990 and 1995 that cited the studies we had previously discovered, and then culled these lists to identify new material relevant to our review.

We used the Dissertation Abstracts International database to identify dissertations relevant to CALL by performing a search on the subject "computer assisted language learning." After inspecting these abstracts, we obtained copies of dissertations that met our criteria from University Microfilms International.

These searches yielded 22 studies that met our six criteria. In order to place these 22 CALL studies within a larger conceptual and historical framework, we also decided to incorporate 13 reviews pertaining to CALL in higher education into our inquiry. We identified these reviews while conducting the earlier searches described above.

Appendix B
Captioning Hardware and Software

Several modestly priced software applications, including "Quick Caption" from The Caption Center and "CPC-600 CaptionMaker" from the Computer Prompting and Captioning Company, make unlimited video captioning in the Roman alphabet possible on a microcomputer for under $3000. With two video-cassette recorders (VCRs), one personal computer, a video monitor, and a "decoder," an individual could add captions in the target language to a video segment by playing the original video on the first VCR, entering one block of text at a time (one to four lines of script prepared ahead of time using a standard word-processing program) by pressing "Enter" on the computer keyboard at the appropriate moment, and recording the captioned video on the second VCR.

"Quick Caption" is available for educational purposes only for about $200 (1995 price) from the Caption Center, WGBH Educational Foundation, 125 Western Avenue, Boston, MA 02134, (617) 492-9225. The CPC-600 CaptionMaker software retails commercially for $1995 (1995 price) from the Computer Prompting & Captioning Company, 1010 Rockville Pike, Rockville, MD 20852-1419, (301) 738-8487. Both applications require the separate purchase of a decoding device, which sells for about $1000. Both applications also assume access to a personal computer with one serial port and one parallel port, two VCRs, and a video monitor.

Captions can also be created in different fonts, font sizes, and colors, as well as in languages such as Russian and Greek, but the software and hardware packages generally cost considerably more. The Computer Prompting & Captioning Company software/hardware configurations with these capabilities, for example, range from about $6000 to $10,000.

As video segments become increasingly frequent components of multimedia CALL, captioning appears to be a worthwhile investment of resources.

Appendix C
Supplementary Findings from the Empirical Studies

For full citations and additional information on the studies cited in this appendix, see Appendix D.

Two studies investigated issues pertaining to cooperative learning and CALL. Kuan-Yi Chang and William Flint Smith (1991) compared the performance of individual students with the performance of pairs of students on the same CALL program, and found no significant achievement differences between students in the two groups. Cristi Mitchell (1992) used random assignment to divide her sample of 55 students into an "individualized" group in which students worked alone and a "cooperative" group in which students worked in groups of 3 or 4. Like Chang and Smith, Mitchell did not find significant achievement differences between students in the two groups.

Three studies examined the role of student learning styles and strategies in CALL. Carol Chapelle and Suesue Mizuno (1989) looked at the learning strategies employed by high-achieving and low-achieving ESL students using the same CALL program, and discovered no overall differences between students in the two groups regarding their use of five basic learning strategies. Richard Raschio (1990) found no significant achievement differences among students with different cognitive styles using the same CALL lesson. Joan Jameison, Leslie Norfleet and Nore Berbisada (1993) examined which factors might help predict student success, failure, or dropout in CALL-type activities, and conclude their article with a profile of a typical student in each of the three categories.

Three other studies focused on computer design issues in CALL. Macarena Aspillaga (1991) investigated whether the physical location of information on a computer screen in a CALL program might be related to student learning. With a sample of 60 students and a research design featuring random assignment, Aspillaga found that students achieved significantly better when the text in a CALL program was superimposed on a graphic or was consistently located in the upper middle section of the screen than when the text was randomly displayed in different sections of the screen. Joan Jameison, Joan Campbell, Leslie Norfleet and Nore Berbisada (1993) investigated how well a computer program could score two types of student writing: student notes and student attempts to recall reading passages they had read earlier. The researchers found a high correlation between the scores given by two human raters and the computer-generated scores, and observed that the human raters averaged about 10 to 15 minutes to score one student's writing, whereas the computer completed the same task in about 1 second. The researchers noted that this finding could be applied to the scoring of open-ended responses in CALL lessons. Ka-Fai Shiu and Sharon Smaldino (1993) compared the effectiveness of CALL and audiotape materials in a Chinese lesson, and found that CALL could be used constructively to teach a non-alphabetic language.

Appendix D

Empirical Studies on CALL in U.S. Higher Education Published Between 1989 and 1994
(Alphabetical Order)

1. Aspillaga, Macarena. (1991). Screen design: Location of information and its effects on learning. *Journal of Computer-Based Instruction*, 18 (3), pp. 89-92.

Sample Size = 60
Language and Level = Beginning Spanish
Time on Computer = One lesson
Randomized Study? = Yes

Research Question

Does (1) "displaying text information overlapped onto relevant parts of a graphic enhance learning" and does (2) "consistency between location of information and pictorial representation facilitate the transfer of presented material into memory?"

Study Design

All students in the study used the same basic CALL program about different parts of a house. During CALL lesson, the Spanish names of rooms and their corresponding English definitions were displayed on the computer screen in different locations. Students were randomly assigned to one of three treatment conditions: (1) text displayed in an area on the computer screen relevant to graphical information, (2) text displayed at upper middle section of screen, and (3) text displayed randomly over a relevant part of the picture, in the upper middle section of screen, or randomly (either the upper middle section of the screen, the bottom middle section of the screen, or overlapping a relevant part of the picture). Each group had 20 students. Software was created by using a NAMING transaction shell.

Sample Information

Sample consisted of 60 undergraduates (from a pool of 360) who had passed the placement test and were now taking entry-level Spanish (Spanish 120).
No native Spanish speakers were in the sample.

Testing

Study used an on-line, built-in test designed by the author with 12 questions and two formats: 1) name of room was displayed and students had to identify the room by using a cursor, and 2) area of house was highlighted and the student had to select the right name of the room (labels for all six rooms were available in multiple choice format).

Results

Group 1 (text on graphic) (mean=10.85) and Group 2 (upper middle section of screen) (mean=11.30). Both scored significantly higher than Group 3 (random) (mean = 7.30). Differences were not significant between mean scores for Groups 1 and 2.

Conclusions

"It is clear that information placed in a consistent location within the monitor screen facilitates the transfer of information and enhancement of learning....Information which is overlapping a relevant aspect of the graphic facilitates transfer of learning, as compared to information placed in random locations" (p. 91).

Reviewer Comments

Information lacking in this short article includes the duration of study, time spent on computer, and the relationship between the researcher and the students in the study.

2. Avent, Joseph. (1993). *A Study of Language Learning Achievement Differences Between Students Using the Traditional Language Laboratory and Students Using Computer-Assisted Language Learning Courseware.* Doctoral Thesis, University of Georgia.

Sample = 272
Language and Level = Intermediate German
Time on Computer = about 6 hours
Randomized Study? = Yes

Research Question

"Are there differences between students who use computer assisted language learning and those students who use the traditional language laboratory? Does the current achievement level of the student [low, medium, high] have an effect on the efficacy of computer assisted language learning?" (p. 7)

Study Design

Students were placed into one of three "achievement level" groups (low, medium, or high) based on their course grades in the previous German course (German 102), and then randomly assigned to the treatment or control group. Students in the control group went to the language laboratory, whereas students in the experimental group went to a Macintosh SE and LC computer lab and worked through the German courseware designed by Avent. Students in the experimental group spent an average of nearly six hours in the computer lab, while students in the control group spent an average of four hours in the language lab (as discussed later, this two-hour discrepancy indicates that students using the CALL program practiced more German, a desirable attribute, but also raises questions of time efficiency). The computer hardware consisted of Macintosh SE and LCs. The courseware covered four units, with each unit including one program focusing on vocabulary and a second program focusing on grammar. Students needed to answer at least 80% of items correct in exercises and achievement checks in order to proceed to next part. Students in both groups attended sections featuring traditional classroom instruction when not in the computer or language labs.

Sample Information

Avent recruited 272 students taking German 103 at the University of Georgia to participate in the study. All students were between the ages of 18 and 25, had successfully completed German 101, and were enrolled in German 102 at the time they volunteered for the study.

Testing

The main instrument was the final exam for German 103, which consisted of a section on listening comprehension and a section on grammar. The final exam allowed direct comparison of achievement between students in the control and experimental groups.

An additional vocabulary test was administered at the end of the quarter to the experimental group alone which included some vocabulary words taught only through CALL and other vocabulary words taught only by traditional methods (e.g.,, oral and written review in the classroom). This vocabulary test permitted comparison within the experimental group of students' understanding of vocabulary words taught either by CALL or by traditional instruction.

Finally, a questionnaire which focused on students' attitudes towards CALL was completed by students in the experimental group.

Results

The mean score on the final exam among students in the CALL group was higher for grammar (Table 1)and vocabulary (Table 2) test items than the mean score of those students who used the traditional language laboratory.

Table 1. Comparison of mean scores of CALL (n=100) and control (n=172) groups on grammar test items on final exam.

	CALL	Control	F	level of significance
mean score	82.2	73.45	49.4	.0001

Table 2. Comparison of mean scores of CALL (n=57) and control (n=119) groups on vocabulary test items on final exam.

	CALL	Control	F	level of significance
mean score	79.7	70	16.4	.0001

For the grammar section, the mean score in each achievement group (low, medium, and high), was also higher for the experimental group than the control group, with significant differences for the "middle group" and for the "high group" (Table 3). The effect size for CALL on the grammar section of the final exam was .680 for the middle group and .611 for the high group.

Similarly, the mean score in each achievement group (low, medium, and high) was also higher for the experimental group than for the control group on the vocabulary test items, with significant differences for the "low group" and the "middle group" (Table 4).

Table 3. Comparison of mean scores of CALL and of control groups on grammar test items on final exam for low (n=42), middle (n=134), and high (n=96) subgroups.

Subgroup	CALL	n	Control	n	t	level of significance
low	66.6	14	60.6	28	1.67	.10
middle	80.7	50	71.4	84	4.07	.0001
high	90.3	36	82.2	60	3.34	.001

Table 4. Comparison of mean scores of CALL and control groups on vocabulary test items on final exam for low (n=33), middle (n=90), and high (n=53) subgroups.

Subgroup	CALL	n	Control	n	t	level of significance
low	76.3	8	57.2	25	2.35	.026
middle	75.5	28	57.2	62	2.21	.029
high	86.5	21	84.8	32	0.53	.596

On the separate vocabulary test, the overall mean scores was higher for those words taught through CALL than through traditional methods (Table 5).

Table 5. Comparison of mean scores of computer group on words taught through CALL (number of words=57) and through traditional methods (number of words=57) on separate vocabulary test.

	words via CALL	words via traditional	F	level of significance
mean score	83.6	74.7	5.48	.0054

Mean scores were also significantly higher for all three achievement groups in the experimental group for words taught by CALL than for words taught through traditional methods (Table 6). The effect size for CALL on the vocabulary section of the final exam was .880 for the low group and .478 for the middle group.

Table 6. Comparison of mean scores of low, middle, and high subgroups of experimental group for words taught through CALL and words taught through traditional methods.

Subgroup	words via CALL	words via traditional	t	level of significance
low	84.1	68.3	5.20	.0013
middle	80.9	71.36	3.34	.0023
high	88.9	82.8	2.26	.037

According to Avent, "in this study it is clear that when the students learned vocabulary items by computerized instruction that, without exception, they remembered them better than the words which had been learned using traditional methods" (pp. 82-83).

Avent also provides questionnaire data from students in CALL group. For example, 84% found "immediate feedback" helpful, 81% reported that they enjoyed their CALL experience in comparison with other types of instruction, and 91% said they would choose CALL again (p. 88).

Conclusions

"The information provided by this study does, it seems, indicate that computer-assisted language learning is effective. It works. Whether or not it is efficient is still somewhat open to question. Regardless of the efficiency or lack thereof, if the goal is for the student to learn the material, then the result of this study would indicate that computer-assisted language learning is a viable alternative, and its development should be pursued" (pp. 96-97)

Reviewer's Comments

This study is noteworthy for its careful design, relatively large sample size, amount of time spent on the computer, and consistently significant results (with some rather large effect sizes). The author's literature review also provides a good overview and introduction to the topic for the generalist reader.

As acknowledged earlier, students in the experimental group on average spent two hours more in the computer lab than students in the control group spent in the language lab. These disparate times pose an intriguing question about the differences in achievement scores at the end of the study between students in the two groups. On the one hand, students using CALL reached higher levels of achievement in German through extended practice in the computer lab. On the other hand, CALL requires more time than traditional instruction in this study. Avent himself addresses this issue in his concluding remarks (cited above) when he asserts that CALL "works" but that its efficiency is "still somewhat open to question."

3. Bationo, Bernadin. (1992). The effects of three feedback forms on learning through a computer-based tutorial. *CALICO Journal*. 10 (1), pp. 45-52.

Sample Size = 56
Language and Level = Beginning French
Time on Computer = 4 lessons
Randomized Study? = Yes

Research Question

"The purpose of this study was to determine which feedback form (written feedback, spoken feedback, or written/spoken feedback) would contribute most to learning intellectual skills in a computer-based language learning tutorial." (p. 47)

Study Design

The computer-based tutorial involved four lessons dealing with the future indicative tense for regular French verbs, and was developed on HyperCard for Macintosh. Participants were randomly assigned to four groups: written and spoken feedback, spoken feedback, written feedback, and no feedback.

Sample Information

The study included 56 undergraduates enrolled in two sections of an Elementary French course at the University of Toledo. Students were told that their participation was voluntary, and were paid $3 at the completion of the study.

Testing

The 14 students in each group took three paper-and-pencil tests: a pre-test, an immediate post-test, and a "delayed" post-test (similar to the immediate post-test) administered two days later.

Conclusions

Very low pre-test scores for all four groups indicated little or no familiarity with the content of the lesson. The only significant difference in average means of post-test results was that of students in the written/spoken feedback group (mean=18.07), who scored higher on immediate post-test than students in the written feedback group (mean=11.5) and in the no feedback group (mean=11.64). The difference between the same written/spoken feedback group (mean=18.07) and the spoken feedback group (mean=14.71) was not significant for the immediate post-test results. No significant results were found among the delayed post-test average means for the four groups (written/spoken feedback mean =15.29; spoken feedback mean =12.14; written feedback mean =10.64; no feedback mean =14.86). The difference between immediate post-test and delayed post-test scores for all groups was not found to be significant.

Reviewer Comments

The no feedback group improved dramatically between immediate and delayed post-tests -- Bationo speculates that "the likely explanation may be that as a result of the frustration, some informal learning took place between the two post-tests which caused high performance scores on the delayed post-test" (p. 50). It is unclear how much time was actually spent on computers during each of the four tutorials, and the "voluntary" nature of study is not entirely clear.

4. Borras, Isabel & Robert Lafayette. (1994). Effects of multi-media courseware subtitling on the speaking performance of college students of French. *The Modern Language Journal*, 78 (1), pp. 61-75.

Sample = 44
Language and Level = Intermediate French
Time on Computer = 2 sessions
Randomized Study? = Yes

Research Question

What are the effects of subtitling (during transactional task practice with multimedia courseware) on the oral communicative performance of fifth-semester college students of French?

Study Design

Students were randomly assigned to one of four treatment conditions:

- subtitled video during oral task practice, lower-level task
- video without subtitles during oral task practice, lower-level task
- subtitled video during oral task practice, higher-level task
- video without subtitles during oral task practice, higher-level task

Students in the "with subtitle" groups had the ability to control the subtitles. Students in all four groups studied two programs (P1 and P2) in a CALL package entitled *Practicing Spoken French*, a multimedia program created by Borras using HyperCard 2.1 and Voyager VideoStack 2.2. Students in the various groups watched a video segment (with or without subtitles) at least two times, and then answered comprehension questions about the video. Next, the students wrote a draft (low-level or high-level) about events they had seen in the video, and finally recorded in French an oral statement up to 3 minutes in length based on their draft. At the end of the experiments, students completed a questionnaire about their attitudes towards the speaking practice sessions and the multimedia courseware package.

Sample Information

Study participants consisted of 44 students enrolled in an intermediate French class at Louisiana State University.

Testing

Oral statements were scored using an assessment instrument developed by Borras that considers effectiveness, accuracy, organization, and fluency.

Results

The students in the subtitle groups scored significantly higher than the students in the control groups on overall oral performance for both CALL programs [P1: $F(1,40)= 74.6$, $p<.001$] [P2: $F(1,40)=68.4$, $p<.001$]. In addition, subscores were higher for the students in the subtitle groups than in the control groups in effectiveness [P1: $F=41.3$, $p<.001$; P2: $F=40.3$, $p<.001$], accuracy [P1: $F=43.1$, $p<.001$; P2: $F=25.9$, $p<.001$], organization [P1: $F=23.3$, $p<.001$; P2: $F=31.7$, $p<.001$], and fluency [P1: $F=33.8$, $p<.001$; P2: $F=33.4$, $p<.001$].

Conclusions

Controlling subtitles can help students learn. Subtitles improve comprehension accuracy, and improve reading literacy.

Reviewer Comments

Borras and Lafayette conclude that study needs to be replicated with larger sample size and different tasks. They also call for further research on the effect of allowing student to control

subtitle insertion and removal and effects of font type and color coding within subtitles. A very similar article by Borras appeared in 1993 in *Educational Technology Research and Development* (#5 below).

5. Borras, Isabel. (1993). Developing and assessing "Practicing Spoken French." *Educational Technology Research and Development*, 41 (4), pp. 91-103.

Basic Information same as Borras and Lafayette, 1994 (see preceding entry)

Results

Mean scores on overall oral performance for the students in the subtitle groups were significantly higher than the mean scores for the students in the control groups. For Program 1, students performing the lower-level task scored higher in the subtitle group (mean=32.3) than students in the group without subtitles (mean=27.1); similarly, on the higher-level task, students in the subtitle group (mean=33.5) outscored the students in the no-subtitle group (mean=28.2). For Program 2, students performing the lower-level task scored higher in the subtitle group (mean=32.9) than students in the group without subtitles (mean=27.7); again, on the higher-level task, students in the subtitle group (mean=34.6) outscored the students in the no-subtitle group (mean=27.7).

Conclusions

"The significance of the results described above point to the effectiveness of a multimedia package that uses computer and videodisk technologies for teaching foreign-language speaking skills" (p. 101).

Reviewer Comments

Borras does not make any explicit concluding statements about subtitling. Discussion at end focuses on *Practicing Speaking French* in general and CALL, not on effects of subtitling. A very similar article co-authored by Borras appeared in 1994 in the *Modern Language Journal* (see #4 above).

6. Chang, Kuan-Yi & Smith, Wm. Flint. (1991). Cooperative learning and CALL/IVD in beginning Spanish: an experiment. *The Modern Language Journal*, 75 (2), pp. 205-211.

Sample = 113
Language and Level = Beginning Spanish
Time on Computer = 2 hours
Randomized Study? = Yes

Research Question

"The purpose of this study was to examine the combined effects of CALL/IVD [Note: IVD stands for interactive videodisk] as a controlled environment used cooperatively for interactive

learning in beginning Spanish when the medium is used to introduce new lesson materials and as a means to assess the relative impact of this type of instruction on learning and the learning environment" (p. 206). In other words, how does the achievement performance of students *working in pairs* on four Spanish lessons using CALL/IVD compare with the performance of individuals *working alone* on the same four lessons using CALL/IVD?

Study Design

Students were randomly assigned to the experimental condition (n=70, students working in pairs in the CALL/IVD workstations) or the control group (n=43, students working individually in the CALL/IVD workstations). The four CALL/IVD lessons averaged between 25 to 35 minutes each, and were developed from *Zarabanda*, an instructional film series produced by the British Broadcasting Corporation. In these lessons, series of multiple-choice questions in Spanish followed segments of video. Error feedback was provided in English, while the correct answer and feedback statement were given in Spanish.

Both treatment and control groups practiced the same four lessons at identical CALL/IVD workstations, with the students in the treatment group working in pairs and the students in the control group working individually. The experiment ran for two weeks (weeks five and six of the semester). The interaction of the students working in pairs was tape recorded, with the participants aware that they were being recorded. Students in pairs were not given any instructions concerning how to interact with their partner.

Sample Information

The sample consisted of 112 freshmen (plus one non-freshman) in ten sections of an accelerated beginning Spanish course at the US Air Force Academy. Students were from diverse geographical locations and socioeconomic backgrounds, and 110 of 113 were between the ages of 18 and 20. All had studied Spanish before for an average of 2 years.

Testing

A 42-item multiple-choice test was administered to all students in the experimental and control groups. which measured the overall comprehension level of each participant on the content of the four CALL/IVD lessons (Cronbach's Alpha=.72). Students in pairs were referred to as the "dyadic" group; students working alone were known as the "monadic" group.

Results

"The results of the t-test performed on the overall achievement scores measured by the achievement test indicated that the mean scores did not differ significantly ($p<.05$) between the two treatment groups due to class structure" (p. 207). On overall achievement scores, the "dyadic" mean was 23.67, close to the "monadic" mean of 23.16 ($t=.46$, $p=.64$). Among the four subcategories of questions--explicit (14 items), implicit (12 items), inferential (13 items), logical-sequential (3 items)--the only category with a significant difference in scores was the implicit measure, where the dyadic students averaged 6.87 and the monadic students averaged 5.89 ($t=2.86$, $p=.005$). Explicit questions: dyadic mean = 6.59, monadic mean = 6.53 ($t=.11$, $p=.91$); inferential questions: dyadic mean = 7.56, monadic mean = 8.16, ($t=-1.24$, $p=.22$); logical-sequential: dyadic mean = 1.8, monadic mean 1.67, $t=.67$, $p=.51$).

Reviewer's Comments
Authors speculate that "the competitive military atmosphere" of the Air Force Academy may have prevented the paired students from engaging in full and open discussion during the CALL/IVD lessons (p. 209). Authors also acknowledge that two weeks may have been insufficient time for students to learn to work constructively in pairs, and that the "dyadic" group might have performed better if they had been given "substantial directions and training in interactive strategies" (p. 209).

7. Chapelle, Carol & Suesue Mizuno. (1989). Students' strategies with learner-controlled CALL. *CALICO Journal*, December issue, pp. 25-47.

Sample = 34
Language and Level = Intermediate English as a Second Language
Time on Computer = 30 minutes
Randomized Study? = No

Research Question
To what extent did "high level" and "low level" ESL students employ five distinct learning strategies when using a CALL program? These five strategies consist of resourcing (using reference materials in the target language), practicing, self-monitoring (correcting own grammar), self-management (creating conditions for own learning), and self-evaluation.

Study Design
The 34 ESL students participating in this study used a CALL program that included grammar lessons designed for intermediate and advanced ESL students. Every keystroke made by each student while engaged in this CALL program was recorded in a separate computer file. The file data were later used to determine the total amount of time each student spent on the grammar lessons, the number of sentences each student constructed, the number of times each student used the help option, how each student edited his or her own work, which phrases each student chose, and which feedback messages each student received from the computer. The authors then used this information to infer how often each student used each of the five learning strategies.

In addition, 13 of the 34 ESL students participating in this study were placed either in a "high proficiency" group (n=7) or a "low proficiency" group (n=6) based on their scores on an English placement test they had taken earlier when entering Iowa State University. Students in the high proficiency group scored greater than one standard deviation above the mean on this test, while students in the low proficiency group scored one deviation or more below the mean.

Sample Information
The 34 students participating in the study were among a pool of 105 students enrolled in five intermediate ESL classes at Iowa State University. These 34 students came from twelve different countries. Ten students were female, and the other 24 were male. All 34 students worked on the CALL grammar lessons for at least 30 minutes and filled out a questionnaire.

Testing

In this study, students did not take an achievement test after completing the CALL grammar programs. Chapelle and Mizuno focused on the process (instead of the product) of student learning through CALL.

Results

Students did not use resourcing strategies often. On average, students requested help once every 8.6 minutes, or once every 4.4 sentences. Students practiced during lessons an average of about 18 minutes, or about 8 sentences.

Differences in the use of all help options between students in the high proficiency group and low proficiency group were small (see Table 1), and significant only for the "requests per sentence" ($p<.05$). Differences in the amount of practice were not significant between students in the high proficiency group and the low proficiency group.

Table 1. Mean number of help requests per minute, number of help requests per sentence, percentage of help requests appropriate for problem-solving, minutes spent on practice, and number of practice sentences for students in high proficiency group (n=7) and students in low proficiency group (n=6).

	high proficiency group	low proficiency group
requests per minute	0.087	0.13
requests per sentence	0.13	0.28
percent of requests/appropriate	71.0	67.0
minutes of practice	15.7	23.5
number of practice sentences	9.8	8.4

Students employed self-monitoring in about 82% of the CALL grammar lessons, self-management in about 81% of the CALL lessons, and self-evaluation in about 46% of the CALL lessons. Differences between students in the high and low proficiency groups in employing these strategies were not considered noteworthy.

Conclusions

Authors emphasize the importance of observing students as they work on CALL lessons and using empirical evidence to design better CALL programs that help students select appropriate learning strategies for particular situations.

Reviewer Comments

Authors do not use a control group in this study. Small sample sizes in the high and low proficiency groups indicate that significance tests should be interpreted with caution.

8. Chun, Dorothy M. (1994). Using computer networking to facilitate the acquisition of interactive competence. *System*, 22 (1), pp. 17-31.

Sample = 15

Language and Level = beginning German
Time on Computer = about 6 hours
Randomized Study? = No

Research Question
Does computer-aided classroom discussion "provide students with the opportunity to generate and initiate different kinds of discourse structures or speech acts"? (p. 20).

Study Design
In this two-semester study, 15 students enrolled in Chun's beginning "honors" German course each engaged in up to 14 real-time class discussions on a local area network, with each discussion lasting about 20-25 minutes. Chun's entire section traveled to the computer laboratory to conduct these on-line discussions in German on topics Chun had announced earlier. During these discussions, participants typed comments and read what others wrote.

Sample Information
During the first semester, 8 women and 6 men were enrolled in Chun's beginning "honors" German course. During the second semester, 8 (4 women and 4 men) of the original 14 students remained, and an additional male student not present during the first semester joined the course.

Testing
No testing was conducted for this study.

Results
Chun found that students averaged 8.4 entries per session, and that the ratio of simple sentences to complex sentences improved from 3 to 1 during the fall semester to 4 to 3 during the spring semester. Virtually every question posed by a student or by Chun during an on-line discussion received an answer, with the total number of replies (229) to Chun's questions numbering about twice as many as the total number of replies (126) to students' questions. The total number of student statements addressed to other students (198), added to the total number of questions asked by students (256), was greater than the total number of replies to questions (454), indicating to Chun that students interacted "directly with each other, as opposed to interacting mainly with the teacher" (p. 28). Chun concludes that the on-line class discussions helped the section move away from the traditional dynamic of teacher-centered interaction in the target language, as students were "definitely taking the initiative, constructing and expanding on topics, and taking a more active role in discourse management than is typically found in classroom discussion" (p. 28).

Reviewer's Comments
Unlike other CALL studies we reviewed involving computer networks, participants in this study did not interact with persons outside of the classroom -- all discussion was with the classmates and the instructor.

9. Cononelos, Terri & Maurizio Olivia. (1993). Using computer networks to enhance foreign language/culture education. *Foreign Language Annals*, 25 (3), pp. 255-267.

Sample = 6
Language and Level = Intermediate/Advanced Italian
Time on Computer = Unclear
Randomized Study? = No

Research Question
How do students enrolled in Intermediate/Advanced Italian respond to the opportunity to communicate with native speakers through UseNET and electronic mail?

Study Design
Students enrolled in an intermediate/advanced Italian course discussed contemporary issues in Italian culture through Usenet (NEWS) and electronic mail (e-mail) with native speakers on the Internet. Students selected a topic of personal interest pertaining to modern Italian culture, such as opera or woman's rights, and investigated the subject through independent study. As of the third week of the course, students posted three messages each week on NEWS, and had to respond to every reply they received at least once. The teacher checked students' contributions to NEWS for the quantity and quality of their writing. Students also responded to messages sent to them through e-mail, but the instructor could not monitor these responses as e-mail accounts are essentially private. At the end of the semester, students turned in a summary and analysis of their postings and the responses elicited by these postings, and evaluated the course in a feedback session.

Sample Information
The six students participating in this study were enrolled in Italian 402-1 ("Topics in Italian Culture: Contemporary Issues"), an intermediate/advanced content-based course at the University of Utah.

Testing
No tests were administered for this study.

Results
Students received an average of three replies for each NEWS posting they wrote. Students reported that they thought their confidence in using Italian and the quality of their writing in Italian improved as a result of this experience.

Conclusions
"The results of Italian 403...suggest that network services are among the CALL tools best suited to content-based, student-centered instruction....The contribution that computer networks make to FL [foreign language] education will ultimately depend on teachers" (pp. 531-32).

Reviewer Comments

Authors acknowledge the concerns of some that potential overuse of NEWS forums in the future by foreign language students could make native speakers less enthusiastic about replying.

10. Fischer, Robert. (1989). Instructional computing in French: the student view. *Foreign Language Annals*, 26 (4), pp. 527-534.

Sample = 34
Language and Level = Beginning French
Time on Computer = about 17 hours
Randomized Study? = Yes

Research Question
How do student perceptions of a CALL program correlate with student achievement? How do students respond to various aspects of a CALL program?

Study Design
Author randomly selected 34 students from second-semester French classes. Students in this experimental group completed 26 drill-and-practice CALL lessons on grammar in place of mechanical lessons on grammar in the writing workbook. Each CALL lesson consisted of 35 items, including vocabulary lessons, irregular verb exercises, discrete-point grammar lessons (verb tenses, object pronouns, comparative structures, and question formation), and integrated grammar exercises (whole sentence translation). Students used CALL in a microcomputer laboratory approximately one hour a week for one semester.

Sample Information
The 34 students in the study were enrolled in second-semester French, apparently at Southwest Texas State University.

Testing
During the last week of the semester, students in the CALL group completed a questionnaire in which they rated various aspects of the CALL program. These students also took an achievement test at the end of the semester that included vocabulary, discrete-point grammar, integrated grammar, and irregular verb morphology.

Results
The only significant correlation between student test scores and student ratings of the usefulness of particular CALL exercises was for vocabulary items ($r = .623$, $p < .001$). Independent of test scores, the questionnaire information also indicated that students thought diagnostic and formal quizzes on CALL were useful, found sound cues to be unhelpful, and accepted drill-and-practice as a valid CALL activity.

Conclusions

"Although much more research is needed in this area, the lack of clear relationships between students' perceptions of these CALL lessons and their relevant posttest scores indicates that they did not generally perceive the instructional value of the lessons directly in terms of their end-of-semester achievement, save for the one area of specific curricular innovation [vocabulary items]" (p. 88).

Reviewer Comments

Author focuses primarily on questionnaire-only data, but lack of correlation among student achievement and student CALL ratings suggests little connection between student perceptions of CALL and student performance on conventional measures. Fischer hypothesizes that the relationship between student achievement and student ratings of vocabulary CALL exercises was significant due to fact vocabulary was not taught during classroom instruction:

> ...the CALL vocabulary lessons provided a unique learning opportunity to the students that was not available to them in normal French classes. The curriculum in the students' classes provided for the teaching and learning of grammatical structures and principles but did not attend to the consistent presentation and practice of vocabulary items. (p. 88)

11. Garza, Thomas. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals*, 24 (3), pp. 239-257.

Sample = 110
Language and Level = Advanced Russian and Advanced English As a Second Language
Time on Computer = Not applicable
Randomized Study? = Yes

Research Question

"What, if any, effect [do] captions themselves have on the learner using captioned [in target language] video materials in the study of a foreign language?" (p. 241).

Study Design

Students were randomly assigned into two groups (with captions and without captions) from class roll lists. Students in both groups attended testing sessions conducted in the same manner. Depending on which group they were in, students viewed an "authentic" video segment with or without captions in the target language. Next, all students were asked to answer ten questions written in the target language for each segment. Students then watched the same video segment again, and had another opportunity to complete a 10-question test. The process was repeated for five video segments in all. Students were instructed to mark only answers for which they had a high degree of certainty, and to leave others blank. At the end of each session, five students were randomly selected to remain for a five-minute individual interview. In the interview, students were asked to retell one video segment of their choosing, keeping as close as possible to the original language of the segment. Interviews were tape recorded. The purpose of the interviews was to determine if captions affect the way advanced students assimilate the inherent language of a video segment.

Sample Information

The sample consisted of 110 undergraduates, with 40 students in the advanced Russian group and 70 students in the advanced ESL group. Russian students were enrolled in 3rd- and 4th-year Russian, and came from the University of Maryland at College Park and Georgetown University. ESL students scored between 500 and 550 on the Test Of English As A Foreign Language (TOEFL) and came from the University of Maryland at College Park and The George Washington University. Russian students were all native English speakers, whereas ESL students as a group spoke 9 different native languages.

Testing

Garza developed the series of 10-question, multiple-choice comprehension checks for each video segment in the target language, which were designed to ensure that the content of the video segments was being tested.

Findings

Students who watched the segments with captions had a mean gain of 75% in correct answers, mean decrease of 61% in incorrect answers, and a mean decrease of 84% in unanswered questions over students who watched without captions. For the ESL students, students in the caption group scored about 30 percentage points higher than control group in correct answers, 22 percentage points lower in incorrect answers, and 8 percentage points lower in unanswered questions (see Table 1). For the Russian students, students in the caption group scored about 38 percentage points higher than control group in correct answers, 17 percentage points lower in incorrect answers, and 21 percentage points lower in unanswered questions (see Table 2).

Table 1. Comparison of overall percentages of correct answers, incorrect answers, and unanswered questions for ESL students in captions group (n=35) and no captions (n=35) group.

	captions	no captions	difference in percentage points
correct answers	87.1%	56.4%	+30.7
incorrect answers	12.3%	34.6%	-22.3
unanswered questions	0.6%	9.0%	-8.4

Table 2. Comparison of overall percentages of correct answers, incorrect answers, and unanswered questions for Russian students in captions group (n=20) and no captions (n=20) group.

	captions	no captions	difference in percentage points
correct answers	82.4%	43.8%	+38.6
incorrect answers	12.6%	30.2%	-17.6
unanswered questions	5.0%	25.9%	-20.9

Average gains in correct responses were higher for Russian students (90%) than for ESL students (60%). Interviewed students consistently demonstrated greater ability to recall the language of the video when they saw captions than students who did not see captions.

Conclusions

"By adding the textual modeling of the captions, the essential language of the segment is made more accessible and, thus, (at least potentially) comprehensible to the learner" (p. 244). "The most significant conclusion suggested by this study is that captioning may help teachers and students of a foreign language bridge the often sizable gap between the development of skills in reading comprehension and listening comprehension, the latter usually lagging significantly behind the former" (p. 246).

Reviewer Comments

This study is exemplary in design and execution. The author notes the difference between the Russian and ESL groups in this study, suggesting the need for caution when generalizing findings across language groups.

12. Hermann, Francoise. (1992). *Instrumental and Agentive Uses of the Computer: Their Role in Learning French as a Foreign Language*. San Francisco: Mellen Research University Press. (Publication of doctoral thesis completed at Stanford University).

Sample = 24
Language and Level = Beginning French
Time on Computer = estimated at 10-15 hours
Randomized Study? = No

Research Question

What are the differences between instrumental and agentive uses of CALL? (p. 21)
(Note: "Instrumental" refers to "using language for action" in socially meaningful tasks; "agentive" refers to "manipulating language," as in drill and practice.)

Study Design

This study involved two sections of a third-quarter French course at the same private university in California. In the first class (n=13), students used CALL to create a classroom newspaper in French. Hermann refers to this group as the "instrumental" group, as CALL is used in an activity-based social context. In the second class (n=11), a different group of students used CALL for drill and practice exercises. Hermann calls this section the "agentive" group. Students were not randomly assigned to these two sections; instead, they enrolled in the different classes on a voluntary and informed basis. Hermann did not teach either section.

Students in the "instrumental" class produced a French-language classroom newspaper using six Xerox d-lion computers at the university's computer center. The different versions of newspaper articles were stored on a shared computer directory that allowed students and the teacher to access all student work on the newspaper in various drafts. Students in this

"instrumental" group also used electronic mail to send messages to each, to their teacher, or to Hermann.

Students in the "agentive" group used an IBM PC computer laboratory to complete a series of nine fill-in-the-missing-word (cloze) sets of exercises in French based on the last eight chapters of the class workbook.

Sample Information

The 24 students participating in the study were enrolled in a third-quarter French course at a private university in California. The sample consisted of 13 females and 11 males. Twenty-two of the students were undergraduates, and 2 were graduate students. These students' prior experience learning French included high school classes and introductory courses on the college level.

Testing

The different versions of articles written by students in the instrumental group, along with their self-selected electronic mail messages, were stored on a computer account. Students in the agentive group stored a record of their computer sessions on individual computer disks. The CALL software for this series of drill and practice cloze exercises provided detailed information on the number and type of correct and incorrect answers for each individual student, which students saved on their disks.

Students in both groups completed a questionnaire about their backgrounds and previous computer use, and students in the instrumental group filled out an additional questionnaire about their CALL experience in producing a newspaper.

Students in both groups completed a battery of four pre- and posttests during the first and last week of their respective courses. This included two standardized tests: the College Board French Listening and Reading Achievement (CEEB) test, and the writing section of the Cooperative Foreign Language (MLA) test (level=MA). Student writing was evaluated on the basis of one-hour compositions, and oral proficiency was assessed through a 15-minute interview with an ACTFL-certified tester.

Results

Differences between mean CEEB test scores for students in the two groups were not significant for the pre-test (instrumental mean=505, agentive mean=520, t=.5), the posttest (instrumental mean=575, agentive mean=570, t=.15), or gains from pre-test to posttest (instrumental mean=70, agentive mean=53, t=.6). [NOTE: Mean gain scores reported by Hermann here and throughout the findings appear to differ from reported pre-test and posttest scores because sample sizes vary slightly in pre-test and posttest conditions. Apparently, all students participating in the study did not complete all four pre- and posttests.]

Differences between mean MLA test scores for students in the two groups were not significant for the pre-test (instrumental mean=34.3, agentive mean=33.4, t=.1) or the posttest (instrumental mean=47.3, agentive mean=53.5, t=1.1). The difference in mean gains, however, between pre-test and post-test was significant (instrumental mean=11.1, agentive mean=20, t=2.2, $p<.05$). [NOTE: Hermann reports instrumental gain as 11.1, even though reported pre-test and post-test scores indicate a 13.9 point difference.]

The writing composition assessment consisted of two assignments: an "imagine" assignment and a "letter" assignment. The mean scores on the writing compositions were not significantly different for either assignment on the pre-test ("imagine": instrumental mean=4.4, agentive mean=5.1, t=.9; "letter": instrumental mean=3.9, agentive mean=4.2, t=.5), the post-test ("imagine": instrumental mean=4.3, agentive mean=5.6, t=1.5; "letter": instrumental mean=4.3, agentive mean=4.8, t=.6), or the gains between pre-test and post-test ("imagine": instrumental mean=.1, agentive mean=.9, t=.7; "letter": instrumental mean=.1, agentive mean=.7, t=.8).

For the pre-test oral proficiency interview, students in the instrumental group scored an average of 3.6 and students in the agentive group scored an average of 3.9. Both scores indicated "intermediate-low" oral proficiency. For the post-test oral proficiency interview, students in the instrumental group scored an average of 5 and students in the agentive group scored an average of 5.3. In this case, both scores indicated "intermediate-middle" oral proficiency.

Conclusions

"The findings of this study indicate that an instrumental approach to the use of the computer in a first year, third quarter French as a foreign language class, and the changes it carries with it, is both an effective and workable alternative approach to L2CALL. Thus, classes in foreign language education could consider using instrumental computer technology in contrast to the prevalent agentive modes of computer use" (p. 159).

Reviewer Comments

Hermann also conducts a series of "T-unit" analyses comparing "an ideal-user case" with "the best-defined instrumental user" in addition to comparing the instrumental and agentive groups. A "T-unit" is basically a main clause plus all the subordinate structures associated with it, and Hermann conducts several comparisons for mean number of T-units and mean length of T-units. Hermann also presents questionnaire, observation, and interview data he collected as part of the study, and discusses the computer data collected from both instrumental and agentive groups. Finally, Hermann includes in Appendix G full text articles from the student newspaper created by students in the instrumental group.

13. Hsu, Jing-Fong, Chapelle, Carol and Ann Thompson. (1993). Exploratory learning environments: what are they and what do students explore? *Journal of Educational Computing Research*, 9(1), pp. 1-15.

Sample = 34
Language and Level = Intermediate/Advanced English as a Second Language
Time on Computer = 4 hours
Randomized Study? = No

Research Question

How do the quantity and quality of student exploration within a CALL program correlate with student attitudes towards computers, learning English, and CALL?

Study Design

The students participating in the study spent one hour each week for four weeks in the computer laboratory learning grammar through a CALL program. For each lesson, students were required to produce ten correct sentences. All keystrokes were recorded for each student in separate computer files over the four CALL sessions. In the CALL program, a student selects a series of short phrases that are placed side by side on the computer screen. Students then must edit the phrases into a complete sentence using the formal rules of English language, including conjugating the verb correctly. The computer provides feedback to the student concerning any problems with the meaning or grammar of each sentence, and the student can then modify the sentence further if desired.

Sample Information

The 34 students participating in the study were all international students at Iowa State University whose writing skills required further development to function successfully in college. The students were enrolled in one of three summer classes: intermediate-level grammar review and composition (n=10); advanced-level composition for undergraduates (n=12); and advanced-level composition for graduates (n=12).

Testing

Students completed a questionnaire at the beginning of the study concerning their attitudes towards computers, learning English, and CALL. Students completed another questionnaire at the end of the study regarding their attitudes towards computers, learning English, CALL in general, and the specific CALL program they used in this study.

Results

Although students had been instructed to construct ten sentences for each session, the number of constructed sentences for each lesson beyond the required ten ranged from 1 to 40, with an average of 16.8. No students experimented further with different grammatical forms after the computer signaled that a constructed sentence was correct. Overall correlations for the whole group of 34 students were not significant between number of completed sentences and attitudes towards computers ($r=.16$), attitudes towards learning English ($r=-.09$), and attitudes towards the specific CALL program ($r=.006$). The one exception was the correlation between number of sentences and attitudes towards CALL in general, which was statistically significant but weak ($r=.25$, $p<.05$).

Correlations were also determined between number of sentences constructed and the four attitude measures for each of three subgroups. Of these twelve correlations, only two were significant, and both involved only the students in the intermediate ESL writing class: the correlation between number of sentences constructed and attitudes towards computers ($r=.56$, $p<.05$), and attitudes towards the specific CALL program used in the study ($r=.56$, $p<.05$). For this intermediate group, the other correlations were between number of sentences constructed and attitudes towards learning English ($r=.50$), and attitude towards CALL in general ($r=.54$).

For the advanced undergraduate ESL writing course, correlations between number of sentences constructed and the various attitude measures were as follows: computers ($r=.003$), learning English ($r=-.42$), CALL in general ($r=.39$), and the specific CALL program ($r=.15$).

For the advanced graduate ESL writing course, correlations between number of sentences constructed and the various attitude measures were as follows: computers (r=.041), learning English (r=-.33), CALL in general (r=-.053), and the specific CALL program (r=-.46).

Conclusions

"It was anticipated that students' attitudes would be correlated with their amounts of exploration. Using number of sentences as the operational definition of exploration, this tendency was seen only for the intermediate group, whereas in the advanced group of graduate students, attitude towards the ESL software was negatively correlated with exploration. We explained these findings by suggesting that the software tended to be at the appropriate level for the majority of intermediate students, while some of the advanced students complained that it was too easy. The latter would have been able to work quickly, producing many sentences, but would have reported a negative attitude towards the ESL software." (p. 13)

"The fact that none of the thirty-four students of varying levels and attitudes used [grammar] exploration indicates that students cannot be expected to employ this strategy on their own. These ESL students, even with positive attitudes toward learning, tended to be more product oriented (focusing on writing correct sentences) than experiment-oriented (focusing on multiple hypotheses) when it came to working with the program." (p. 13)

Reviewer Comments

Authors hypothesize that students did not creatively explore the grammar capabilities of the CALL program in part because they did not receive explicit instruction concerning how they might investigate this aspect of the program.

14. Jameison, Joan, Campbell, Joan, Norfleet, Leslie & Berbisada, Nora. (1993). Reliability of a computerized scoring routine for an open-ended task. *System*, 21(3), pp. 305-322.

Sample = 40
Language and Level = Not applicable
Time on Computer = estimated at 5 hours
Randomized Study? = No

Research Question

"Can students' notes and recalls of reading passages be scored by a computer program as reliably as they are by people?" (p. 307).

Study Design

Nine teachers of thirteen sections of freshman composition classes at Northern Arizona University volunteered to participate in this study. The research team then visited each of these thirteen classes to recruit student volunteers for the study. Students interested in participating in the study were told to use the computer laboratory at their convenience during a four-month period. Participants studied reading, note-taking, and recall through four computerized lessons on music appreciation, minerals, electric conductivity, and Aztec civilization. In each lesson, students read

one paragraph at a time in the upper half of the computer screen and typed notes in the space provided on the bottom half of the screen. At least one day later, students used their notes to type as much as they could remember about the lesson.

The 40 students chosen for the data analysis from the total sample of 286 were not randomly selected. Rather, ten students were selected for each of the four lessons because these ten had recently finished that particular lesson. The mean age and distribution of gender and first language reflected that of the larger sample of 286 students.

The hardware used in this study was a mainframe DEC VAX computer, and the lessons were authored using Digital's Courseware Authoring System (CAS), Digital Authoring Language (DAL), and "C."

Sample Information

The 40 students participating in the study were enrolled in a freshman composition class at Northern Arizona University. The students ranged in age from 18 to 37, with an average of 20. The sample consisted of 15 males and 25 females. The first languages of students included English (n=36), Navajo (n=2), and Spanish (n=1).

Testing

Students' notes for the music and electric conductivity lessons and recalls for the minerals and Aztecs lessons were scored by hand by two of the authors. Raters evaluated the number of discrete idea units (IUs) present in the original lessons that were also present in each set of notes, with each IU assigned a predetermined value. An idea unit is the smallest unit of information that contains a unique assertion and can thus be judged true or false. Raters also identified the number of "superordinate" (more important) and "subordinate" (less important) IUs.

A computer program was also developed by the authors to score the same sets of notes and recalls.

Results

Table 1. Comparison of mean scores for student notes on a music and an electricity lesson rated by human scorers (n=2) and a computer on different "information unit" (IU) dimensions.

Music lesson

	human	computer
total IUs	46.9	53.1
superordinate IUs	5.1	5.7
subordinate IUs	11.4	12.9

Electricity lesson

	human	computer
total IUs	56.2	52.5
superordinate IUs	3.5	5.4
subordinate IUs	32.2	28.2

The reliability coefficient between the two human scorers was quite high overall on the two sets of notes (r=.98, r=1.0) and the two sets of recalls (r=.99, r=.89). Reliability was also high between the average of the human scores and the computer program on the two sets of notes (r=.91, r=.94) and the two sets of recalls (r=.96, r=.93).

For the two sets of notes, mean human and computer scores were similar (see Table 1). For the two sets of recall, mean human and computer scores were also similar (see Table 2).

Table 2. Comparison of mean scores for student recall on a minerals and an Aztecs lesson rated by human scorers (n=2) and a computer on different "information unit" (IU) dimensions.

Minerals lesson

	human	computer
total IUs	20.1	20.2
superordinate IUs	3.1	1.5
subordinate IUs	9.2	10.8

Aztecs lesson

	human	computer
total IUs	14.5	16.7
superordinate IUs	2.4	2.2
subordinate IUs	7.2	8.8

The human scorers took about 10 to 15 minutes to score one student's notes or recall, while the computer program took about 1 second to perform the same task.

Conclusions

"Our research question was answered in the affirmative: students' notes and recalls of reading passages can be scored by a computer program as reliably as they are by people. Overall, the computer program scored as the human scorers did" (p. 316).

"The technique described here presents a viable alternative for measuring open-ended responses. If we want to move toward language tasks that are more like real life, the open-ended response is one avenue that needs to be further explored" (p. 318).

Reviewer Comments

While this is not a CALL study in the traditional sense, the first author is a well-recognized CALL researcher and the abstract explicitly links this study to CALL: "this article asserts the value of open-ended responses for CALL lessons and language tests" (p. 305). The authors believe this study demonstrates the practicality of developing assessment capabilities within CALL programs that move beyond multiple-choice questions to open-ended responses scored by computer.

This article also contains a detailed description and explanation of the computer program used to score the notes and recall.

15. Jamieson, Joan, Norfleet, Leslie & Berbisada, Nora. (1993). *Successes, failures, and dropouts in computer-assisted language learning*. ERIC ED 354 786.

Sample = 158
Language and Level = Beginning English Composition
Time on Computer = Unclear
Randomized Study? = No

Research Question
Which factors--among individual characteristics, strategies, and course information-- help predict student success, failure, or dropout in CALL-type activities?

Study Design
No randomization in this study. The authors looked at naturally occurring results, and then attempted to identify factors which might have predicted those outcomes. Students were supposed to complete 4 computerized reading and note-taking lessons on expository prose. Computerized lessons were presented on a VAX computer system, and programmed in Digital's Authoring Language and "C." After each lesson, students completed 2 achievement tests and an attitude questionnaire. At the end of the testing period, the authors divided students into three groups: successful (scored above 70th percentile on all 8 achievement measures, n=41), failure (scored below 30th percentile on all 8 measures, n=27), and dropouts (n=90).

Sample Information
Study involved 158 students enrolled in freshman composition classes at Northern Arizona University.

Testing
Before students started computerized lessons, data was collected on students' age, sex, first language, second language, and time between end of high school and beginning of college. Students also took the Group Embedded Figures Test (GEFT) to measure field independence/dependence (a dimension of cognitive style). These data were used to analyze factors pertaining to "individual characteristics." In addition, students completed a questionnaire--which was an adapted version of the Strategy Inventory for Language Learning--at the same time that they took the GEFT. The questionnaire helped the authors determine whether students were likely to employ direct or indirect strategies.
During the testing sessions, students took a recall test and a 20-question multiple-choice comprehension quiz, both of which were conducted on the computer.

Results
"The successful student was one who had a high semester and cumulative GPA, took a high number of units both during the semester and cumulatively, and had a high Field Independence/Dependence score. On the other hand, a student who belonged to the Failure group was one who had a low semester and cumulative GPA, had taken fewer semester and

cumulative units, and had a low Field Independence/Dependence score. There was a complexity in the student profile of the Dropout group because though he had a low semester and cumulative GPA and few semester and cumulative units, his Field Independence/Dependence score was only a little bit lower than the Success group average, but much higher than that of the Failure group." (p. 14)

Conclusions

The authors cite the need for individualized instruction on CALL. "We must not forget that [computers] are to assist us and that the adaptation need not be on the part of our CALL participants, but rather the adaptation can be on the part of the computer" (p. 15).

Reviewer's Comments

This study is not with a foreign language class, but an English composition course. The authors consider the study to be a research and development CALL project on study skills. The authors' central focus is on role of student characteristics in CALL. Students were not assigned randomly to groups, and the study is a post-hoc analysis of naturalistic outcomes. The authors also provide some stepwise regression results.

16. Mitchell, Cristi. (1992). *The Relationship of Computer-Assisted Language Learning Environments and Cognitive Style to Achievement in English as a Second Language.* Doctoral Thesis, University of Miami.

Sample = 55
Language and Level = Intermediate English as a Second Language
Time on Computer = 7 hours
Randomized Study? = Yes

Research Question

"Are there differences in achievement between adult ESL students working individually and ESL students working in cooperative CAI environments? Are there differences in achievement between auditory second language learners and visual second language learners working on CAI lessons? Is there an interaction between the CAI learning environments and learning style for adult ESL students?" (pp. 5-6)

Study Design

Mitchell randomly designated the two sections as the individualized CALL group (n=31) and the cooperative CALL group (n=24). Students in both groups completed a series of 15 CALL tutorial and drill-and-practice assignments on the past tense in English in a Macintosh and Apple IIe computer laboratory over a period of 3 weeks. Students in the individual treatment group used CALL by themselves, whereas students in the cooperative group were placed into groups of 3-4 students using stratified random assignment (with each group including a low-, middle-, and high-scorer on the pre-test).

Sample Information

The 55 students participating in the study were enrolled in two different sections of intermediate (level four) ESL at the Miami-Dade Community College South (Kendall) Campus.

Testing

Participants took the Hill Learning Styles Survey during the first week of class to identify either an auditory or a visual learning style. A teacher-designed pre-test and post-test both consisted of 45 multiple-choice questions on simple past and perfect tenses.

Results

No significant differences were found between the individual and cooperative groups for the variables of gender, native language, learning style, pre-test scores, absences, or tardies, suggesting initial equivalence between the two groups.

In the individual group, 16 students were identified as primarily auditory learners, 13 as primarily visual learners, and 2 equally auditory and visual. In the cooperative group, 13 students were identified as primarily auditory learners, and 11 as primarily visual learners.

Both groups scored significantly higher on the post-test than on the pre-test (individual: pre-test mean=55.2, post-test mean=72.4; cooperative: pre-test mean=56.4, post-test mean=74.8; combined: $t=5.67$, $p<.001$). Differences between mean scores of the two groups on the post-test, however, were not significant for treatment group (individual vs. cooperative) ($F=.05$, $p=.83$), learning style (auditory vs. visual) ($F=.21$, $p=.65$), or interaction of treatment group and learning style ($F=.45$, $p=.51$).

Conclusions

"Since the ESL students in the cooperative learning environments did not show significantly greater achievement when provided with training, practice, and assignments in cooperative CAI than students working individually on CAI assignments, there is need for further investigation of the factors influencing the effectiveness of cooperative CALL environments" (p. 71).

"Although the results of this study revealed that differences in student achievement were not impacted by the preferred learning styles of ESL students, it is difficult from these results to clearly establish the relationship between learning styles and achievement. Several possible factors including sample size and procedure, selection of classes, assessment instruments, and definitions of learning style should be considered" (p. 74).

Reviewer Comments

Random assignment took place on the level of class, not on the level of individual.

17. Nagata, Noriko. (1993). Intelligent computer feedback for second language instruction. *The Modern Language Journal*, 77 (3), pp. 330-339.

Sample = 34
Language and Level = Intermediate Japanese
Time On Computer = 4 hours
Randomized Study? = Yes

Research Question

"The study compared two versions of the Nihongo-CALI exercises, the traditional CALI (T-CALI) exercises (involving conventional feedback) and the intelligent CALI (I-CALI) exercises (providing sophisticated feedback), and investigated whether the differences in the amount and quality of feedback, based on error analysis, would affect the learners' performance on tests of production of Japanese passive sentences....The study also sought to determine whether significant differences exist in the learners' attitudes toward T-CALI and I-CALI exercises" (p. 335). [Note: "traditional" CALI (computer-assisted language instruction) error feedback provides information about what the student did wrong, whereas "intelligent" CALI gives detailed explanation of why student response was wrong.]

Study Design

Participants were paired on the basis of a written test that assessed knowledge of basic Japanese grammar, and then randomly assigned to the I-CALI or T-CALI group. Students engaged in four one-hour sessions with the computer, and "were not aware that different types of feedback were being compared" (p. 336). Students took an achievement test shortly after completing the fourth computer session, and a retention test three weeks after the end of the experiment. The Nihongo-CALL software used in the I-CALI group incorporated an artificial intelligence approach called Natural Language Processing, which ran on both Macintoshes (minimum 5 MB of RAM) and DEC workstations.

Sample Information

Study participants consisted of 34 students in a second-year Japanese language course at the University of Pittsburgh.

Testing

The achievement test consisted of 20 questions. Three weeks later, the participants took a final exam in which four questions related to passive structures, the subject of the CALI unit.

Results

The achievement test results (perfect test score=34) showed a significant difference between the average mean scores of the I-CALI (mean=27.9) and T-CALI (mean=26.9) groups ($t=2.18$, $p<.05$). Again, a significant difference was found on the "passive structure" component of the final exam (perfect test score=6.8) between the average mean scores of the I-CALI (mean=6.0) and the T-CALI (mean=5.3) groups ($t=2.34$, $p<.05$). During the fourth computer session, participants were asked to complete a 28-item, 5-point scale to rate different aspects of the software programs. Students in the I-CALI group rated the error messages much higher than students in the T-CALI group.

Reviewer Comments

Nagata states "the results suggest that the traditional feedback may be as good as the intelligent feedback for helping learners to correct word-level errors (e.g., vocabulary and conjugation errors), while the intelligent feedback may be more helpful for understanding and correcting sentence-level errors (e.g., particle errors), which involve more complex processing of

knowledge" (p. 337). Nagata also acknowledges that "without careful study of what feedback to provide, intelligent CALI may yield an overflow of useless information" (p. 338). Though the sample size is small, the statistically significant findings are noteworthy, and the conclusion that CALL with intelligent feedback is particularly well-suited for sentence-level errors suggests the effectiveness of CALL with relatively complex language issues.

18. Nieves, Kelly. (1994). *The Development of a Technology-Based Class in Beginning Spanish: Experiences with using* EXITO. Doctoral Thesis, George Mason University.

Sample = 37
Language and Level = Beginning Spanish
Time on Computer = estimated at 50-60 hours
Randomized Study? = No

Research Question
How well do students learn Spanish using multimedia CALL compared to a traditional college classroom? (pp. 86-87).

Study Design
Nieves converted *EXITO*, a multimedia CALL program in Spanish originally developed by the Central Intelligence Agency as a 10-day/60-hour intensive course, into a one-semester college course in introductory Spanish at George Mason University, and then conducted a formative evaluation of this course under development. Nieves inverted the typical ratio of computer time to classroom time in this experimental course, as students spent four to five hours per week in the computer lab and only one hour per week in class with the instructor.

In addition to the largely qualitative formative evaluation of EXITO, Nieves included quantitative data comparing the performance of 19 students in the CALL group with another 18 students in a control group of students taking beginning Spanish without a CALL component.

Nieves also analyzed interviews conducted by a faculty member with 8 students in the CALL group and 9 students in the control group about their experiences in their Spanish class.

Sample Information
The 37 students participating in this study were enrolled in a first-semester beginning Spanish course at George Mason University.

Testing
Nieves developed a Spanish proficiency test designed to assess skills in listening, speaking, reading, and writing. The test consisted of multiple-choice question for the listening and reading sections, and open-ended questions for the speaking and writing sections.

Results
Nieves found that students in the CALL group had a higher mean score (mean=97) than the control group (mean=90) on the Spanish proficiency exam (maximum score=160), and a much

smaller range (range=65) than students in the control group (range=112). Further, when Nieves broke down these mean scores by "true beginners" and "false beginners" (with the former representing students who had never studied Spanish before), the difference between the EXITO and control groups became more pronounced. The mean score on the proficiency exam was substantially higher for "true beginners" in the CALL group (mean=84) than for "true beginners" in the control group (mean=60).

Conclusions

"The technology-based SPAN 101 course that the researcher developed for this dissertation project offers an effective alternative to the traditional textbook approach to teaching Spanish" (p. 129).

"The EXITO students who were not confident in their skills in Spanish were all 'true beginners' who had never studied Spanish before this semester....Still, the 'true beginners using EXITO underestimated their abilities in Spanish. They were able to out-perform the 'true beginners' in the control class. Apparently, 'true beginners' who are left to work on their own with material in a new target language are not as sure of their progress because they do not get as much encouragement and reinforcement from the teacher" (p. 129).

Reviewer Comments

The primary focus of Nieves' dissertation is not the quantitative data in this pilot study. Rather, this quantitative component is situated within Nieves' broader research about her conversion of EXITO into a one-semester college course and her qualitative formative evaluation of the course.

19. Raschio, Richard. (1990). The role of cognitive style in improving computer/assisted language learning. *Hispania*, 73 (May), pp. 535-541.

Sample = 62
Language and Level = Beginning Spanish
Time on Computer = 2 class periods
Randomized Study? = Yes

Research Question

What role does cognitive style play in computer assisted language learning?

Study Design

Author randomly assigned students into an experimental group (n=33) and a control group (n=29). Students in the experimental group learned about pronouns for indirect and direct objects through a CALL tutorial, whereas students in the control group learned the same lesson through print material. Both the CALL tutorial and print materials were prepared by Raschio. The instructor was available to render assistance in the computer laboratory. After two class periods dedicated to this lesson (either through CALL or the print material), students in both groups were tested on the subject matter.

Tests
Achievement tests used in study were written by Raschio.

Results
No achievement differences were found between the two groups. No relationship was found between "field dependence" and student achievement level. Furthermore, no relationship was found between student attitude and either student achievement or field dependence. "Field independent" learners had the most comments and asked for the most assistance in the computer lab, whereas "field central" learners had fewest suggestions.

Conclusions
Raschio emphasizes the need to understand students' learning strategies in order to design CALL lessons that take advantage of these individual styles.

Reviewer Comments
Raschio does not report t-statistics or p-values in this article.

20. Shiu, Ka-Fai & Sharon Smaldino. (1993). *A pilot study: Comparing the use of computer-based instruction materials and audio-tape materials in practicing Chinese.* ERIC ED 362 204.

Sample = Not stated
Language and Level = Beginning Mandarin Chinese
Time on Computer = Not stated
Randomized Study? = No

Research Question
"In comparing audio-tape vs. CALL, does either medium led to better learning of Mandarin Chinese?"

Study Design
All participants in the study were enrolled in an intensive Mandarin Chinese class in which audio-tape and CALL lessons were used on alternate weeks throughout the semester-long course. The authors created the CALL lessons using HyperCard on Macintosh.

Sample Information
Students in the study were enrolled in a beginning intensive Mandarin Chinese course at the University of Northern Iowa.

Tests
Weekly achievement tests written by the authors were administered to the students in the study.

Results

The authors found no difference in listening comprehension tasks, but found that students performed better on translation and character writing tasks when those lessons were taught through CALL.

Conclusions
"It is possible to use CBI [computer-based instruction] materials to teach non-alphabetic languages."

Reviewer Comments
The authors do not provide data or details in this paper to support their findings.

21. Stenson, Nancy, Downing, Bruce, Smith, Jan & Smith, Karin. (1992). The effectiveness of computer-assisted pronunciation training. *CALICO Journal*, Summer issue, pp. 5-19.

Sample = 36
Language and Level = Advanced English as a Second Language
Time on Computer = about 80 minutes
Randomized Study? = No

Research Question
"The specific hypothesis tested in this study was that ITAs [international teaching assistants] using SpeechViewer [an IBM software program which provides visual representations of speech] would make greater progress in their overall pronunciation and, in particular, their stress, rhythm, and intonation, as well as in their ability to pronounce key words in their academic fields, than would those ITAs working with more traditional methods of pronunciation practice." (p. 7)

Study Design
Students in experimental and control groups attended one two-hour group session each week, with 4 students assigned to each group session. Each student also received 50 minutes of one-on-one instruction every week. Students in treatment groups had instructors who used Speech Viewer regularly in the one-on-one tutorials, while students in the control group did not use SpeechViewer at all during their 50-minute sessions. An average session on SpeechViewer lasted 15 minutes during a 50-minute tutorial session, while individual times varied from 5 to 42 minutes (duration was not recorded in 11 of the 88 sessions with SpeechViewer). Of the eight tutorial sessions during study, students in treatment groups used SpeechViewer an average of 5.5 sessions. Average total amount of time with SpeechViewer for students in treatment group is 80 minutes (assuming average values for missing 11 entries). Five of 13 "ITA" instructors used SpeechViewers with their ITA students. Full pre- and post-tests were available for 18 students (of the original 25) in the experimental group. "These 18 were matched with an equal number of ITAs from among the 35 members of the control group. ITAs in the experimental and control groups were matched as closely as possible for the following (in descending order of importance):

scores on the pre-course SPEAK test, native language, and academic discipline" (p. 9). The IBM SpeechViewer operated on an IBM PS-2 Model 30 with a 30-megabyte hard drive.

Sample Information

Students were enrolled in this quarter-long course at the University of Minnesota because they did not reach the minimum score on the SPEAK Test (Educational Testing Service) required for working as a teaching assistant at the University of Minnesota. In addition to ESL instruction, this course developed teaching skills appropriate for working effectively as a teaching assistant, and ITAs were videotaped presenting short lessons which focused on a particular teaching skill.

Testing

The authors of this study used the SPEAK Test, available through the Educational Testing Service, and the Mimic Test, a test of English language designed for the study by the researchers. In the Mimic Test, students were asked to listen to a native speaker pronounce words, phrases, and sentences, and then repeat them, mimicking the model as closely as possible. Mimic was designed "to provide a maximally constrained opportunity for ITAs to attempt native-like pronunciation in a setting where content was not a concern, in order to compare their success on this task with their success in more natural, communicative settings" (p. 17) "Although reliability and validity have not been established for the Mimic Test and our results were mixed, we note that other researchers have relied on similar tests" (p. 17).

Results

Despite claims of general widespread enthusiasm for SpeechViewer, no significant differences were found between pre- and post-test scores for treatment and control groups on both SPEAK and Mimic tests. (SPEAK Pre-test: Treatment average score = 188.41, Control = 187.60; SPEAK Post-test: Treatment = 209.85, Control 208.56)

Conclusions

"While these results may mean that SpeechViewer does not have a significantly greater effect than traditional methods on the pronunciation skills of ITAs (as measured by these tests), it is also possible that the ITAs simply did not get enough practice with SpeechViewer to show dramatic results." (p. 13) "The fact that the quantitative results do not show more than very minor differences between the experimental and control groups, while the qualitative results suggest that instructors and ITAs alike were enthusiastic about the use of SpeechViewer, is problematic." (p. 14).

Reviewer's Comments

Authors also collected qualitative information during the study. Instructors using SpeechViewer during their tutorials recorded comments in a logbook, and these entries indicated much enthusiasm for the program by instructors and ITAs. Table 1 appears to be incomplete and inaccurate. While the table presents SPEAK pre- and post-test scores for experimental and control groups, the table only provides Mimic pre- and post-test scores for experimental group (scores for control group missing). Furthermore, the post-test score for the experimental group

(Mimic2) is inconsistent with the text of the article. The article states "both groups improved their performance on the Mimic test, the experimental group by 7.33 points..." but the numbers in the table indicate the experimental group's performance on Mimic actually declined by 19.67 points. Most probable explanation: a typographical error occurred, and 236.15 should read 263.15.

Authors suggested that SpeechViewer might be used for self-instruction (p. 13). Authors also mentioned that several issues emerged with the SpeechViewer software. Sometimes a "perfectly acceptable" performance by an ITA did not correspond to the "correct" model on SpeechViewer (p. 15). To some extent, the differences in pitch ranges between male and female voices caused these discrepancies. Authors also stated that saving student work in the form of digitized speech required considerable hard-drive storage capacity--the 30MB hard drive used in this study proved inadequate, and recorded material was lost.

22. Wright, David Allan. (1992). *The reciprocal nature of universal grammar and language learning strategies in computer assisted language learning*. Masters Thesis, University of Arizona.

Sample = 107
Language and Level = Beginning German
Time on Computer = Unclear
Randomized Study? = No

Research Question

"The first purpose of this study was to determine whether a difference exists in German 102 [second semester German] students' achievement on chapter examinations depending on whether they used the *Auf Deutsch* computerized workbook or the standard *Auf Deutsch* workbook. The second purpose was to determine whether a difference exists in the students' understanding of language learning strategies based on the Strategy Inventory for Language Learning." (p. 46).

Study Design

Experimental group (n=45) consisted of students in three sections of German who used computerized workbooks for vocabulary and grammar study. Control group consisted of students from three other sections of same course who used standard workbooks. Students in experimental sections still used standard workbooks for listening and communication exercises. One experimental section was randomly chosen from a group of seven sections, whereas the other two experimental sections were taught by Wright (the author). Students in experimental sections had access to computers at three different sites from 8 a.m. to 11 p.m. 7 days a week.

Computerized workbooks and standard workbooks provided similar content and exercises, but computerized workbooks were also able to give instant feedback and suggestions for finding correct answers. Computerized workbooks could also help explain why an answer was correct and list the page number in the textbook where an explanation could be found, while the standard workbook only provided an answer key to the questions at the back of the book without explanation.

Sample Information

Study participants consisted of 107 students in six sections of German 102 at the University of Arizona.

Testing

Students took three unit tests. First, mean vocabulary scores were compared between experimental and control groups for each unit. Second, mean overall test scores were compared between experimental and control groups for each unit. Third, mean combined overall test scores (for the three units together) scores were compared between experimental and control groups. A two-sample t-test was used to analyze results.

Results

For vocabulary scores, only the results on the Chapter #1 exam were significantly different (see Table 1). For unit exams, the results for Chapter Exam #1 and Chapter Exam #2 were significantly different (see Table 2). For total test scores, the experimental group mean scores were significantly higher than the experimental group (CALL mean=85.5, CNTRL mean=81.3; t=3.17, p=.0017).

Table 1. Comparison of mean scores of CALL and control groups on vocabulary items on three chapter exams (n=number of students).

	CALL	n	Control	n	t	level of significance
Chapter #1	12.2	52	9.86	44	4.94	.00001
Chapter #2	9.6	43	8.66	53	1.21	.23
Chapter #3	8.3	48	7.75	64	1.01	.31

Table 2. Comparison of mean scores of CALL and control groups on three chapter exams (n=number of students).

	CALL	n	Control	n	t	level of significance
Chapter #1	84.9	49	79.6	59	2.25	.027
Chapter #2	85.3	48	80.4	62	2.37	.02
Chapter #3	86.4	49	85.0	42	0.55	.59

This study also provided results for student responses to the Student Inventory for Language Learning (SILL) on learning strategies for experimental and control groups. Students in the experimental group scored higher for three of the five strategies on SILL: "I read a story or dialogue several times until I can understand it," "I revise what I write in the new language to improve my writing," and "I use reference materials such as glossaries or dictionaries to help me use the new language" (p. 70).

Conclusions

"The computer is a flexible classroom aid which can be used by researchers, teachers and learners in a variety of ways and for a variety of purposes. By presenting this discussion, it has been

demonstrated that CALL must be applied within the framework of current SLA [second language acquisition] theory so that it can become an even more useful tool in the learner-centered, proficiency-oriented classroom" (p. 73).

"The classroom use of CALL needs to be determined by teachers and not by software writers. Teachers know most about the needs and capacities of their students, and about how the learning experiences should be structured" (pp. 73-74).

Reviewer Comments

Wright noted that this study coincided with the pilot project that was later implemented into regular German 101/102 curriculum. Since the author was involved with the pilot project and taught two of the three experimental sections, "it was impossible to eliminate teacher/researcher bias completely" (p. 55). As random assignment was done on level of section (not by individual) for only one of the three sections in experimental group, internal validity concerns arise concerning whether results be attributed to CALL or the effect of the teacher/researcher on the performance of the experimental sections.

Appendix E

Reviews of CALL and CAI, 1987-1992, by Category:
Syntheses of Reviews, Narrative Reviews, Focused Reviews, and Meta-Analyses

A. Syntheses of Reviews

1. Dunkel, Patricia. (1991). The effectiveness research on computer-assisted instruction and computer-assisted language learning. In *Computer-Assisted Language Learning and Testing: Research Issues and Practice*. Dunkel, Patricia, editor. New York: Newbury House. pp. 5-36.

Number of References: 98.

Purpose of the Review: To determine the impact of computers on learning, particularly language learning.

Findings:

1. Students master subject matter content faster with CAI.
2. Studies of student attitude changes as a result of computer use show mixed results.
3. CAI seems more effective in science, math, and languages than in other subjects.
4. CAI effectiveness differs by age of student and type of computer use.
5. CAI is more effective as supplement to, not replacement for the instructor.

Recommendations:

1. Researchers need to move from asking "does CALL work" to asking what kinds of CALL lessons work with what kinds of learners under what kinds of conditions.
2. Researchers need to investigate the social as well as the cognitive impact of using computers for language learning and teaching.
3. Researchers need to focus on questions of instructional design variables in CALL lessons, such as screen design, feedback and branching of the CALL program, student cognitive approaches to various types of CALL lessons, student learning styles and CALL lessons, student control of the program, and cost-effectiveness of CALL.
4. Language teachers must become intimately involved in the CALL research process.

2. Niemiec, Richard, and Walberg, Herbert J. (1987). Comparative effects of computer-assisted instruction: a synthesis of reviews. *Journal of Educational Computing Research*, 3, (1), pp. 19-37.

Number of Reviews Reviewed: 16, representing over 250 studies.

Number of References: 24 additional references.

Purpose of the Review: Niemiec and Walberg describe 16 reviews of CAI, examine the methodology of each, synthesize the conclusions of the reviews, and suggest directions for future research.

Further Comment:

1. For inclusion in this synthesis, the review had to meet 3 criteria:
   a. examine studies which looked at CAI and achievement and/or affect measures.
   b. include a minimum of 4 primary studies occurring in actual classrooms.
   c. not be a republication of an earlier review.
2. For each review, the following was reported:
   a. search and selection procedure
   b. analysis methodology
   c. instructional level
   d. type of computer application
   e. outcomes measured
   f. number of studies included
   g. effect size (calculated by Niemiec and Walberg)

Findings:

1. CAI is effective in improving student achievement.
2. Effect sizes vary with type of computer use and students characteristics.

Recommendations:

Researchers need to address several questions:

1. What kinds of instructional designs work well with what kinds of students?
2. Are linear or branching programs more effective? If branching, what type of branching is more effective?
3. Is CAI cost effective?
4. How does CAI compare with other instructional media with regard to efficacy and cost effectiveness?

3. Roblyer, M.D., Castine, W.H., and King, F.J. (1988). *Assessing the Impact of Computer-Based Instruction*. New York: The Haworth Press.

Number of Reviews Reviewed: 26

Number of References: 41

Purpose of the Review: To discover trends in the field of CAI and to suggest directions for future research.

Findings:

1. Students learn faster with computers.

2. Although student attitude seems positive with CAI, motivation for learning seems unaffected.
3. CAI seems effective across content areas.
4. Efficacy of types of CAI seems to depend on a number of variables, such as the instructional strategy of the lesson and the nature of the skill being taught.
5. The computer is more effective as a supplement to the work of the teacher, rather than as a replacement.
6. CAI seems to work with all ability levels, and seems to work better when used with the target population for which the specific lesson was designed.
7. CAI, CMI, and CEI seem to be differentially effective by grade level. These findings are summarized in Tables 11 and 12 in the body of *On CALL* (pp. 37-38).

Recommendations:

1. Because a number of variables determine the outcome of any comparison of CAI with traditional teacher instruction, a researcher cannot simply compare effect sizes across studies to determine the *best* kinds of computer lessons. Researchers need to identify and describe all the varibles in the study, such as student characteristics, and the instructional design of both the computer instruction and traditional instruction.
2. As the field of CAI software design matures and teachers begin using the improved lessons, additional reviews of the ongoing CAI research will be necessary to determine what software designs work well with which students.
3. Researchers must include an analysis of the cost-effectiveness of CAI as compared to traditional instruction.
4. Effect sizes for CAI lessons should be compared to other alternative strategies to traditional instruction, such as cooperative learning or peer tutoring.
5. Future studies and reviews should emphasize microcomputers rather than mainframes.
6. Future meta-analyses should perform separate calculations of effect sizes for different content areas and target groups.

4. Williams, Carol J., and Brown, Scott W. (1991). A review of the research for use of computer-related technologies for instruction: An agenda for research. In *Educational Media and Technology Yearbook - 1991*. Branyan-Broadbent, Brenda, and Wood, R. Kent, editors. Englewood, CO: Libraries Unlimited. pp. 26-46.

Number of References: 71

Purpose of the Review: To summarize what has been learned from the research comparing CAI with conventional instruction, to identify issues for research in instructional technology, and to suggest directions for future research.

Findings:

1. CAI improves student achievement.
2. Problems exist in the design and reporting of studies of CAI effectiveness.

Recommendations:

Future CAI research needs to:

1. Focus on how learner characteristics interact with CAI to yield differential outcomes.
2. Be designed from a foundation of learning theory.
3. Ask what kinds of CAI work for what kinds of learners and for what kinds of learning.
4. Use the power of computers to record students' keystrokes as a research tool.

B. Narrative Reviews

5. Chapelle, Carol, and Jamieson, Joan. (1989). Research trends in computer-assisted language learning. In *Teaching Languages With Computers: The State of the Art*. Pennington, Martha, editor. La Jolla, CA: Athelstan. pp. 45-60.

Number of Studies Reviewed: 29

Number of References: 42

Purpose of the Review: To determine trends in CALL research with regard to student achievement and attitude, and to describe interactions of various CALL strategies with student characteristics.

Findings:

1. In control-treatment designs, where the computer is used by the treatment group, neither CAI nor CALL can claim unequivocal superiority over traditional instruction.
2. Comparisons of lesson strategies with CALL in individual studies suggest that some lesson strategies are more effective than others with some students in some circumstances.
3. Studies examining the interaction between lesson strategies and learner characteristics such as learning style (e.g., field independence), find that for some types of students, certain types of lesson strategy are more effective.
4. When working in groups, students exhibit different amounts and different kinds of language interaction with each other when engaged in different kinds of CALL lessons.

NOTE: The findings reported in 2, 3, and 4, are supported by too few studies to generalize the conclusions.

Recommendations:

1. Future studies need to reflect the principles of the theories of cognitive psychology and second language acquisition when examining aspects of lesson design and the particular learners for whom the lessons are being designed.

2. Future studies need to account for learning style differences among students and how these differences interact with different CALL lesson strategies.
3. Collection of on-line data as students are working on CALL programs is necessary to analyze the factors influencing the learning outcomes for each student, and to determine what learning strategies are used by students in response to lesson characteristics.
4. CALL programs should be analyzed along seven dimensions:
    a. Type of activity (e.g., drill, game, tutorial)
    b. Type of learning (e.g., grammar, vocabulary, global language use)
    c. What the student is actually doing when using the CALL lesson
    d. Program focus (the linguistic purpose of the lesson)
    e. Level of language lesson (e.g., beginner, expert)
    f. Program difficulty (software flexibility with regard to the learner's performance)
    g. How the CALL lesson is integrated into the language class as a whole

6. Garrett, Nina. (1991). Technology in the service of learning: Trends and issues. *Modern Language Journal*, 75, (1), pp. 74-96.

Number of References: 12

Purpose of the Review: To provide an overview of the technology available to support language learning and to determine the efficacy of this technology.

Findings: The research base cannot at this time support an unequivocal claim that CALL is effective.

Recommendations:

1. CALL research needs to look closely at individual variables affecting lesson outcomes. Specifically, the CALL researcher needs to examine what kinds of CALL software work with what kinds of students for what kinds of desired learning outcomes.
2. Experienced language teachers who are also experienced in CALL and who keep abreast of computer hardware developments are in the best position to make decisions about the purchase of CALL software and hardware. However, few language teachers have such expertise.
3. CALL studies need to be designed within the theoretical framework of modern second language acquisition theory, which suggests that language learning is acquiring the ability to communicate in the new language.
4. Studies need to determine what kind of error analysis and feedback works with what kind of students.

7. Pederson, Kathleen M. (1987). Research on CALL. In *Modern Media in Foreign Language Education: Theory and Implementation*. Smith, Wm. Flint, editor. Lincolnwood, IL: National Textbook Company. pp. 99-131.

Number of References: 92

Purpose of the Review: To evaluate basic and applied research in CALL, and to offer a theoretical base for further research.

Findings:

1. Few studies have reported what specific pedagogical manipulations in CALL software are effective with what students and with what lessons.
2. The financial and human resources to conduct CALL research are scarce.
3. The results of a national survey of workers in the field of language teaching suggested a need for more and better software and for more and better teacher training in the use of CALL.

Recommendations:

1. CALL research design must take into account all the specific variables associated with the lesson strategy (the computer's coding options), desired learning outcomes, and student characteristics.
2. CALL researchers must provide a theoretical basis for their research design, and must carefully operationalize each of the variables in the study.
3. Pederson suggests that CALL researchers use Gavriel Salomon's theory for designing media research because it accounts for the three sets of variables affecting learning, namely learner variables, the coding elements of the medium, and the specifics of the learning task. The five tenets of Salomon's theory that are important in the design of CALL research are as follows:
    a. The medium's coding elements are what affect learning, not the medium per se.
    b. Coding elements are most effective when they activate cognitive subskills such as focusing, organizing, or highlighting.
    c. Coding elements affect different learners in different ways.
    d. Learners differ in their perceptions of the task expectations communicated by various coding elements.
    e. Complex interactions result from the interplay of coding elements, learning tasks, and learner differences

8. Scott, Tony, Cole, Michael, and Engle, Martin. (1992). Computers and Education: A Cultural Constructivist Perspective. In *Review of Research in Education - 18*. Grant, Gerald, editor. Washington, DC: American Educational Research Association. pp. 191-254.

Number of References: 142

Purpose of the Review: To examine the pedagogical use of computers in K-12 education and to identify trends most relevant to actual classroom practice.

Findings:

From the Apple Classroom of Tomorrow Project:

1. Students took the initiative for their own academic work.
2. Students wrote more and wrote better.
3. Students spontaneously worked collaboratively.
4. Teachers became coaches rather than direct instructors.
5. Students learned basic skills faster.
6. Extensive teacher training is necessary to effectively use a computer-rich environment for teaching.

Other findings:

1. CAI programs interact with student gender, ethnicity, and first language.
2. Computer networks within schools and among schools result in a decompartmentalization of the learning experience for students.

Recommendations:

Practitioners and others who initiate the use of computers in the classroom need to design carefully crafted evaluation strategies in order to discover how to best use the technology with each student.

9. Smith, Wm. Flint. (1988). Modern media in foreign language education: A synopsis. In *Modern Media in Foreign Language Education: Theory and Implementation*. Smith, Wm. Flint, editor. Lincolnwood, IL: National Textbook Company. pp. 1-12.

Number of References: 11

Purpose of the Review: To provide a broad overview of the state of CALL research in 1987.

Findings:

1. CALL implementation is hampered both by poor availability of teacher training and little demand by language teachers for such training in CALL.
2. Published CALL materials vary widely in quality.
3. Many CALL programs are authored by language teachers with little computer programming expertise or by programmers with little knowledge of language acquisition theory and language teaching pedagogy.

Recommendations:

Training needs to be made available to language teachers in the use of CALL, and in CALL authoring languages and computer programming techniques.

C. Focused Reviews

10. Chapelle, Carol, and Jamieson, Joan. (1991). Internal and external validity issues in research on CALL effectiveness. In *Computer-Assisted Language Learning and Testing: Research Issues and Practice.* Dunkel, Patricia, editor. New York: Newbury House. pp. 37-60.

Number of References: 65

Purpose of the Review: To examine the variability of studies in the CALL effectiveness research base, and to offer guidelines for future research that takes into account validity issues.

Findings:

1. In quasi-experimental CALL research, variables not specifically accounted for, such as learner characteristics or the context of the CALL lesson, threaten the validity of the CALL study.
2. In descriptive research examining student attitudes toward CALL, contraints of time and funding prevent the researcher from assuring that all covariates to student attitude have been accounted for.
3. The generalizeability of a CALL study is limited unless sufficient detail is reported about student characteristics, the classroom context, and the characteristics of the CALL lesson.
4. CALL research as of 1991 cannot support consistent and unambiguous statements about what kinds of CALL lessons work with what kinds of students in what contexts.

Recommendations:

CALL researchers need to pay attention to the validity of their studies by doing the following:

1. In quasi-experimental studies, all variables need to be operationalized and accounted for.
2. In studies about student attitudes, care must be taken to assure candid student responses.
3. Descriptive studies must be consistent in definition and application of concepts of students learning strategies, linguistic functions, and other variables under consideration.
4. Researchers need to report on the variables they were unable to control or account for, and which, as a result, may threaten the validity of their study.

11. Johnson, Donna M. (1991). Second language and content learning with computers: Research in the role of social factors. In *Computer-Assisted Language Learning and Testing: Research Issues and Practice.* Dunkel, Patricia, editor. New York: Newbury House. pp. 61-84.

Number of References: 73

Purpose of the Review: To examine the research on social aspects of student computer use in classrooms.

Findings:

1. Simply creating opportunities for students to interact in the target language does not guarantee the students will respond.
2. Teachers need to carefully plan student interactions around the computer lesson, especially in designing the goal structure for the lesson.
3. Rich use of the target language results from pairs of students collaborating in a word-processed composition in the target language.
4. Students who communicate in the target language through email networks exhibit high motivation and rich use of the target language.

Recommendations:

Teachers need to plan carefully their use of computers in language teaching so that the kinds of student interaction stimulated result in acquisition of the target language.

D. Meta-analyses

12. Roblyer, M.D., Castine, W.H., and King, F.J. (1988). *Assessing the Impact of Computer-Based Instruction*. New York: The Haworth Press.

Number of Studies Reviewed: 38 studies, 44 dissertations

Number of References: 116

Purpose of the Meta-analysis: The review of reviews in this same volume by Roblyer et al. (summarized above, Syntheses of Reviews, 3) reviewed research done primarily on mainframe computers. The research on microcomputer use by teachers in the 80's had not been widely reviewed at the time of their review of reviews. Therefore they decided to perform a meta-analysis of the available research on classroom use of microcomputers.

Further comment:

Inclusion criteria: Studies were selected that

1. were done between 1980 and 1987.
2. are free from major methodological flaws.
3. have control groups.
4. reported group sizes, means, standard deviations.
5. used microcomputers (with a few exceptions).

Findings:

1. Of the studies and dissertations analyzed, most were done in the last three years of the 1980-1987 time period.

2. Most of the studies were in math or reading, most were in basic skills, and about half were in elementary grades.
3. Student attitudes toward school and subject matter were more positive among students using the computer.
4. The small number of studies precludes drawing any definitive conclusions with regard to differential effects of computer-based education (CBE) on subject matter. However, it appears that teaching cognitive skills such as problem-solving and critical thinking yields about the same effect sizes as for reading and math.
5. With regard to type of application of CBE (e.g., tutorial, simulation, drill and practice), results varied across subject areas. A sufficient number of studies to support such an analysis was found only reading and math. Various types of math applications seemed equally effective. Tutorial applications in reading yielded larger effect sizes that other types of computer applications. These differences, however, were not statistically significant at $p<.05$. In science, simulations yield higher effect sizes than for drill and practice.
6. With regard to age of students, CBE can be successful at all ages.
7. With regard to types of students, software designed for specific groups of students (e.g., slow learners) seems to be more effective with the target population than with other students. Based on only a few studies, CBE may not be as effective with students whose first language is Spanish.

Recommendations:

1. Too few studies have been reported to support generalizations regarding trends. Therefore, research in CBE must be increased. Roblyer et al. suggest three ways in which this may be done, as follows.
    a. Educational organizations must accept the notion that research is necessary for improvement of student learning, and therefore provide the necessary funding and organizational support.
    b. Funding agencies should solicit research projects in CBE designed to answer key questions.
    c. Practitioners must insist that their organizations support their keeping abreast of and implementing research in instructional computing.
2. More studies are needed to determine what types of computer applications are more effective in which content areas.
3. Evaluations are needed to determine the effectiveness of computer applications for English as a second language (ESL), especially for Spanish-speaking students.
4. More work needs to be done to determine how to use word processing programs to improve the quality of students' writing.
5. More studies are needed to determine what kinds of software programs designed to teach problem-solving and critical thinking skills are more effective with what kinds of students.
6. More work is needed to determine the relationship among CBE, student attitudes toward school and learning, and dropout rates.

7. Studies are needed to determine whether various CBE programs have a differential effect on female vs. male students.
8. Cost effectiveness of CBE needs to be compared with that of other educational interventions, such as cross-age tutoring, increasing instructional time, or reducing class size.
9. With regard to standards for research reports, Roblyer et al. note that many of the studies considered for their meta-analysis provided insufficient detail on study design and/or on their statistical procedures, or failed to report descriptive data, and hence could not be included in the meta-analysis. They suggest, therefore, that researchers follow the *Publication Manual* of the American Psychological Association (1983), reporting at least statistical data, sample sizes, test instrument information, treatment information, and design information.